# TRANSACTIONS

## OF THE

# AMERICAN MATHEMATICAL SOCIETY




VOLUME 52

JULY TO DECEMBER, 1942

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1942

Chester C. Corbin Library Fund
Bound January, 1943

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

# TABLE OF CONTENTS

## VOLUME 52, JULY TO DECEMBER, 1942

# QUADRATIC DIOPHANTINE EQUATIONS IN THE RATIONAL AND QUADRATIC FIELDS

BY

IVAN NIVEN

**1. Introduction and summary.** It is well known that the equation

$$lx + my + n = 0, \tag{1}$$

with rational integral coefficients, has either no solution in rational integers or an infinite number of solutions. The same result is true in quadratic fields, that is, when $l$, $m$ and $n$ are integers of a given quadratic field, and solutions are sought among the integers of the field.

We are here concerned with the number of integral solutions of the general quadratic equation

$$ax^2 + bxy + cy^2 + dx + ey + f = 0, \qquad a \neq 0; A = b^2 - 4ac, \tag{2}$$

with integral coefficients from the field of rational numbers or from some quadratic field. The quantity $A$ is defined for convenient reference. We can take $a \neq 0$ without any loss of generality, by the use (if necessary) of linear transformations of determinant unity (so that the number of integral solutions is not changed).

First, suppose that the coefficients of (2) are rational integers. If $A$ is negative, then the graph of (2) is finite in extent, and there is at most a finite number of solutions in integers. If $A \geqq 0$, the graph of (2) is a parabola, an hyperbola, or two straight lines, and we prove the following result.

THEOREM 1. *Let the coefficients of equation* (2) *be rational integers, with* $A \geqq 0$. *Then if* (2) *has one solution in integers, it has an infinite number, with the following exceptions: if* (2) *represents two essentially irrational straight lines, it has at most one integral solution; if* (2) *is an hyperbola whose asymptotes are essentially rational, then it has at most a finite number of integral solutions.*

By an *essentially rational* straight line, we mean one whose equation can be put in the form (1), with rational integral coefficients; otherwise we say that a line is *essentially irrational.*

Next, suppose that the coefficients of (2) are from a real quadratic field. It turns out in this case that we can have an infinite number of integral solutions when the curve is finite in extent. Also, the hyperbola does not divide into two cases, as it does in Theorem 1. Before stating the theorem, we recall the definition that a totally negative quadratic integer is a negative integer

Presented to the Society, September 5, 1941; received by the editors April 29, 1941.

whose conjugate is also negative. Thus $-5-2^{1/2}$ is totally negative, whereas $-5-4\ (2^{1/2})$ is negative but not totally negative.

THEOREM 2. *Let the coefficients of* (2) *be integers of a real quadratic field F. Then if* (2) *has one solution in integers of F, it has an infinite number, except in the following cases: if* (2) *represents a point, or a pair of straight lines whose coefficients are essentially outside the field F, then it has at most one integral solution in F; if* (2) *represents an ellipse (so that A is negative), and A is totally negative, then it has at most a finite number of integral solutions in F.*

Finally, suppose the coefficients of (2) are from an imaginary quadratic field. Our result is much the same, but there are interesting differences.

THEOREM 3. *Let the coefficients of* (2) *be integers of an imaginary quadratic field F. Then one solution of* (2) *in integers implies an infinite number of such solutions, with the following exceptions: if the left side of* (2) *factors into two linear expressions in x and y, with coefficients essentially outside F, then* (2) *has at most one integral solution in F; if $A\neq 0$ is the square of an integer of F, and the left side of* (2) *is not factorable into linear expressions in x and y, then* (2) *has at most a finite number of integral solutions in F.*

In proving Theorems 2 and 3, we use the Pell equation in quadratic fields,

$$\xi^2-\gamma\eta^2=1. \tag{3}$$

In this connection, we prove the following theorem.

THEOREM 4. *Let $\gamma$ be an integer, not zero, of a quadratic field F. Then equation* (3) *has an infinite number of integral solutions $(\xi,\eta)$ in the field F if and only if $\gamma$ is not the square of an integer of F when F is imaginary, and $\gamma$ is not totally negative when F is real.*

That the conditions of this theorem are sufficient to insure an infinite number of solutions of (3), is proved in the next two sections. The necessity of the conditions follows from Theorems 2 and 3, as we shall see at the end of §3.

Theorems 1, 2, and 3 are sufficiently similar that the principal results can be proved by a common method; this is presented in §4. Then the theorems are completed in the last three sections. The methods employed throughout the paper are elementary.

2. **The Pell equation in quadratic fields.** If $\gamma$ is a positive rational integer, not a square, it is well known that equation (3) has an infinite number of solutions in rational integers. We can obtain a similar result for quadratic fields from a classical theorem on the units of relatively cyclic fields.

Let $\gamma$ now be an integer of a quadratic field $F$. If $\gamma$ is not a perfect square in $F$, and not a rational integer, then the biquadratic field $K=R(\gamma^{1/2})$ is rela-

tively cyclic of prime order two over $F$. It is known(1) that there exists a relative unit of norm 1 in the field $K$ over $F$, provided that among the four conjugate fields determined by $K$ there are twice as many real fields as there are among the two conjugate fields determined by $F$. This condition is satisfied when $F$ is an imaginary field, because the conjugate of $F$, being identical with $F$, is also imaginary; consequently, equation (3) has a non-trivial integral solution (that is, a solution with $\eta \neq 0$) in $F$ provided $\gamma$ is not a perfect square in $F$.

On the other hand, if $F$ is a real quadratic field, then it is again identical with its conjugate, and we require $K$ and its conjugates to be real. Now $K$ and its conjugates are identical in pairs, each field being either $R(\gamma^{1/2})$ or $R(\bar{\gamma}^{1/2})$, where $\bar{\gamma}$ is the conjugate of $\gamma$ in $F$. These are real provided that $\gamma$ and $\bar{\gamma}$ are positive, or in other words, provided that $\gamma$ is totally positive. Hence we can conclude that if $\gamma$ is a totally positive integer of a real quadratic field $F$, then equation (3) has a non-trivial integral solution in $F$.

Having one solution of (3), we can obtain more by means of the composition formula

$$(\xi_1^2 - \gamma\eta_1^2)(\xi_2^2 - \gamma\eta_2^2) = (\xi_1\xi_2 + \gamma\eta_1\eta_2)^2 - \gamma(\xi_1\eta_2 + \xi_2\eta_1)^2.$$

This provides an infinitude of different solutions. For example, a non-trivial solution compounds with itself to give a different non-trivial solution. We have proved this lemma.

LEMMA 1. *Let $\gamma$ be an integer, not a square, of any quadratic field $F$. Let $\gamma$ be totally positive if $F$ is real. Then equation* (3) *has an infinite number of integral solutions in $F$.*

3. **Real quadratic fields.** Lemma 1 is not the best possible result for real quadratic fields. We now prove:

LEMMA 2. *If $\gamma$ is a positive, but not totally positive, integer of a real quadratic field $F$, then equation* (3) *has an infinite number of integral solutions in $F$.*

We prove this by a method analogous to that of Dirichlet(2) for rational and Gaussian integers. Let $F$ be obtained by extending the rational numbers by $m^{1/2}$, $m$ being positive, square-free, and greater than 1. Then $\gamma$ has the form $a+bm^{1/2}$, where $a-bm^{1/2}$ is negative. For convenience, let $\delta$ denote the positive square root of $\gamma$. For any positive rational integer $n$, we let $v$ range over the values $1, 2, \cdots, n+1$. Let $u$ be the greatest integer less than $vm^{1/2}$, that is, $u=[vm^{1/2}]$, and we have

(1) Cf. David Hilbert, *Die Theorie der algebraischen Zahlkörper*, Jahresbericht der Deutschen Mathematiker-Vereinigung, vol. 4, p. 275 (Theorem 92) and p. 279.
(2) Cf. Dickson, *History of the Theory of Numbers*, vol. 2, p. 373.

$$|u - vm^{1/2}| < 1. \tag{4}$$

Choose $y$ and $x$ as follows:

$$y = [\delta(u + vm^{1/2})/(2m^{1/2})], \qquad x = [\delta(u + vm^{1/2}) - ym^{1/2}] + 1.$$

These equations imply the inequalities

$$0 \leqq \delta(u + vm^{1/2}) - 2ym^{1/2} < 2m^{1/2},$$

and

$$0 < x + ym^{1/2} - \delta(u + vm^{1/2}) \leqq 1, \tag{5}$$

respectively, and these add to give the result

$$0 < x - ym^{1/2} < 1 + 2m^{1/2}. \tag{6}$$

As $v$ ranges over the integers $1, 2, \cdots, n+1$ the expression involved in (5) takes values between 0 and 1, at least two of which differ by less than $1/n$. We subtract these to obtain

$$X + Ym^{1/2} - \delta(U + Vm^{1/2}) < \frac{1}{n}, \tag{7}$$

and the inequalities (4) and (6) imply that the rational integers $X$, $Y$, $U$ and $V$ satisfy

$$|U - Vm^{1/2}| < 2, \qquad |X - Ym^{1/2}| < 1 + 2m^{1/2}. \tag{8}$$

Using the fact that $|V| \leqq n$, we can write

$$\begin{aligned} | X + Ym^{1/2} &+ \delta(U + Vm^{1/2}) | \\ &\leqq | X + Ym^{1/2} - \delta(U + Vm^{1/2}) | + 2 | \delta(U + Vm^{1/2}) | \\ &< \frac{1}{n} + 2\delta | U - Vm^{1/2} | + 2\delta | 2Vm^{1/2} | \\ &< \frac{1}{n} + 2\delta(2 + 2nm^{1/2}). \end{aligned}$$

The multiplication of this inequality by (7) gives

$$| (X + Ym^{1/2})^2 - \delta^2(U + Vm^{1/2})^2 | < 1 + 2\delta(2 + 2m^{1/2}),$$

and we set $\xi = X + Ym^{1/2}$ and $\eta = U + Vm^{1/2}$ to obtain

$$| \xi^2 - \gamma\eta^2 | < 1 + 2\delta(2 + 2m^{1/2}). \tag{9}$$

We now show that this inequality is satisfied by an infinitude of pairs $(\xi, \eta)$. The left side of inequality (7) is not zero, for otherwise $\delta$ would be an element of the field $F$. Then $\gamma$, being the square of an element of $F$, would be totally positive, contrary to hypothesis. Now if the number of pairs of quad-

ratic integers satisfying (9) were finite, the rational integer $n$ could be chosen so large that none of these pairs would satisfy (7). Our method would therefore give another pair of values satisfying (7) and (9).

Having shown that (9) represents an infinite number of inequalities, we now show that $\xi^2-\gamma\eta^2$ assumes only a finite set of values. We cannot conclude this directly from inequality (9), because there are infinitely many integers of a real quadratic field which are less in absolute value than a given positive quantity. However, there is but a finite number of integers of such a field which, *together with their conjugates*, are bounded in absolute value. Since $\bar{\gamma}$ is negative, we can use (8) to obtain

$$\left|\bar{\xi}^2 - \bar{\gamma}\bar{\eta}^2\right| = (X - Ym^{1/2})^2 - \bar{\gamma}(U - Vm^{1/2})^2 < (1 + 2m^{1/2})^2 - \bar{\gamma}(4).$$

Hence the infinite set of quadratic integers $\xi^2-\gamma\eta^2$ of inequality (9) ranges over a finite set of values. At least one of these values, say $\rho$, is equal to $\xi^2-\gamma\eta^2$ for an infinite number of pairs

$$(10) \qquad (\xi_1, \eta_1), (\xi_2, \eta_2), \cdots .$$

We now show that it is possible to select from (10) an infinite subsequence

$$(11) \qquad (\xi_{i_1}, \eta_{i_1}), (\xi_{i_2}, \eta_{i_2}), \cdots ,$$

such that

$$(12) \qquad \begin{aligned} \xi_{i_j} - \xi_{i_k} &\equiv 0 \pmod{\rho}, \\ \eta_{i_j} - \eta_{i_k} &\equiv 0 \pmod{\rho}. \end{aligned} \qquad (j, k = 1, 2, 3, \cdots),$$

Let the quantities $\xi_1, \xi_2, \cdots$ of (10) be written as

$$(13) \qquad X_1 + Y_1m^{1/2}, X_2 + Y_2m^{1/2}, \cdots ,$$

the $X_i$ and $Y_i$ being rational integers. Let $N(\rho)$ denote the norm of $\rho$. Since each $X_i$ $(i=1, 2, \cdots)$ is congruent to some term of the complete residue system $0, 1, 2, \cdots, N(\rho)-1$, modulo $N(\rho)$, it follows that an infinite number of these are congruent to one particular term of this residue system. Thus from (13) we have selected an infinite subsequence, and from the latter we can select another so that the $Y_j$ are congruent to one another modulo $N(\rho)$. We continue this process of selecting subsequences with the terms $\eta_i$ of (10), and obtain finally a sequence (11) such that congruences analogous to (12) hold modulo $N(\rho)$, and these imply (12).

We now select two different pairs from (11), say $(\xi_r, \eta_r)$ and $(\xi_s, \eta_s)$; let these be independent in the sense that one pair is not the negative of the other pair. They satisfy the relations

$$(14) \qquad \xi_r^2 - \gamma\eta_r^2 = \xi_s^2 - \gamma\eta_s^2 = \rho,$$

and we multiply these equations to get

$$(15) \qquad (\xi_r\xi_s - \gamma\eta_r\eta_s)^2 - \gamma(\xi_r\eta_s - \xi_s\eta_r)^2 = \rho^2.$$

But the congruences (12) imply that the integer $\xi_r\eta_s - \xi_s\eta_r$ is divisible by $\rho$. Consequently $\xi_r\xi_s - \gamma\eta_r\eta_s$ is divisible by $\rho$, and we obtain a solution of (3) by dividing (15) by $\rho^2$. The solution thus obtained is not trivial, that is, $\xi_r\eta_s - \xi_s\eta_r \neq 0$. For otherwise we could write $\xi_r = k\xi_s$ and $\eta_r = k\eta_s$ with $k \neq \pm 1$, and these relationships contradict (14). Noting that an infinite number of solutions of (3) can now be obtained by the method set forth at the end of §2, we have completed the proof of the lemma.

LEMMA 3. *Let $\gamma$ be a negative, but not totally negative, integer of a real quadratic field $F$. Then equation* (3) *has infinitely many integral solutions in $F$.*

This is a direct consequence of Lemma 2. For, by hypothesis the integer $\bar{\gamma}$ is positive. Hence there are infinitely many solutions of

$$\xi^2 - \bar{\gamma}\eta^2 = 1.$$

The conjugates of these solutions are solutions of (3), and the lemma is proved.

LEMMA 4. *Suppose that $\gamma = \alpha^2 \neq 0$, where $\alpha$ is an integer of a real quadratic field $F$. Then equation* (3) *has an infinite number of integral solutions in $F$.*

As in Lemma 2, we take $F$ to be $R(m^{1/2})$. When $\alpha$ is multiplied by its conjugate $\bar{\alpha}$, the result is a rational integer, the norm of $\alpha$, say $n$. Now there are infinitely many pairs of rational integers satisfying

$$u^2 - mn^2v^2 = 1,$$

since $mn^2$ is not a square. Taking $\xi = u$ and $\eta = m^{1/2}\bar{\alpha}v$, we obtain infinitely many solutions of (3).

Lemmas 1, 2, 3, and 4 give all cases of Pell equations (3) in quadratic fields with an infinite number of solutions, for it is a consequence of Theorems 2 and 3 that equation (3) can have but a finite number of integral solutions for values of $\gamma$ other than those stated in the above lemmas. Thus, upon proving these theorems, we shall have Theorem 4 as a consequence.

4. **The general theory.** We return our attention to equation (2), the coefficient field $F$ being the rational numbers or some quadratic field. Solving for $x$ we get

$$(16) \qquad x = \frac{1}{2a}\left\{-by - d \pm (Ay^2 + By + C)^{1/2}\right\},$$

where $B = 2bd - 4ae$ and $C = d^2 - 4af$.

*Case* 1. $B^2 - 4AC \neq 0$; $A$ is positive and not a square when $F$ is the field of rational numbers; $A$ is neither zero nor totally negative when $F$ is a real quad-

ratic field; $A$ is not the square of an integer of $F$ when $F$ is an imaginary quadratic field. With these hypotheses, we show that one integral solution of (2) implies an infinite number. Let there be such a solution $(x_0, y_0)$. Then there exists an integer $t_0$ such that the equation

$$t^2 = Ay^2 + By + C \tag{17}$$

is satisfied by the values $t_0$, $y_0$. We substitute these values in (17), and subtract the result from (17), to get an equation which can be written in the form

$$(t - t_0)(t + t_0) = (y - y_0)(Ay + Ay_0 + B). \tag{18}$$

We look for integral solutions of this equation. We write

$$p(y - y_0) = 2aq(t + t_0), \quad 2aq(Ay + Ay_0 + B) = p(t - t_0), \tag{19}$$

where $p$ and $q$ will be specified later. Eliminating $t$ from these equations, we get

$$(p^2 - 4a^2Aq^2)y = p^2y_0 + 4a^2q^2(Ay_0 + B) + 4pqat_0. \tag{20}$$

By the hypotheses of the case under discussion, and by the lemmas of the last two sections, we can choose the integers $p$ and $q$ in infinitely many ways so that

$$p^2 - 4a^2Aq^2 = 1. \tag{21}$$

Thus we obtain integral values for $y$ in (20). These, in turn, give integral values for $t$ in (19), as can be seen by eliminating $y$ from these equations.

We now make certain that these values of $y$ give integral values of $x$ in (16). Multiplying the first equation in (19) by $p$, and eliminating $p^2$ by the use of (21), we see that

$$(y - y_0) + 4a^2A(y - y_0) = 2apq(t + t_0).$$

Hence $y \equiv y_0 \pmod{2a}$, and the same argument applied to the second equation in (19) shows that $t \equiv t_0 \pmod{2a}$. These imply the congruence

$$-by - d \pm t \equiv -by_0 - d \pm t_0 \pmod{2a}.$$

Since $y_0$ and $t_0$ give the integral value $x_0$ in (16), this congruence shows that our method gives integral values for $x$, provided that the sign is chosen properly.

Finally we must demonstrate that the above procedure gives an infinitude of solutions of (16). Using (21) to eliminate $p^2$ from (20), we have

$$y = y_0 + 4aq\{aq(2Ay_0 + B) + pt_0\}. \tag{22}$$

First, suppose that $t_0 = 0$. Then $2Ay_0 + B \neq 0$, for otherwise we could write $y_0 = -B/2A$, and these values of $t_0$ and $y_0$, when substituted in (17), give $B^2 - 4AC = 0$, contrary to hypothesis. Also $a \neq 0$, so that the coefficient of $q^2$

in (22) is not zero. Consequently, each different value of $q^2$ gives a different value of $y$.

In the second place, if $t_0 \neq 0$, we show that of all the values satisfying (21), only a finite number give $y = y_0$ in (22). Values of $p$ and $q$ giving $y = y_0$ satisfy $aq(2Ay_0 + B) + pt_0 = 0$, and the result of eliminating $p$ from (21) by means of this equation is

$$q^2\{(2aAy_0 + aB)^2 - 4a^2At_0^2\} = t_0^2.$$

This is satisfied by not more than two values of $q$.

Suppose now that equation (22) gives only a finite set of values, say $y_0, y_1, \cdots, y_r$. We select a rational prime $\pi$ which does not divide any of $y_1 - y_0, y_2 - y_0, \cdots, y_r - y_0$. Let $(P, Q)$ be such a solution of

$$P^2 - 4a^2A\pi^2Q^2 = 1$$

that the corresponding solutions $p = P$, $q = \pi Q$ of (21) do not give $y = y_0$ in (22). Then the value $y$ thus obtained from (22), having the property that $y - y_0$ is divisible by $\pi$, is different from $y_1, y_2, \cdots, y_r$. We have shown, therefore, that (22) gives an infinite set of different values.

*Case* 2. $A = 0$, $B^2 - 4AC \neq 0$, so that $B \neq 0$. Again we assume one integral solution $(x_0, y_0)$ of (16), and show that it can be used to generate an infinite number. Proceeding as we did in the first case, we get the following equation analogous to (18)

$$B(y - y_0) = (t - t_0)(t + t_0).$$

We write

$$y - y_0 = 2aq(t + t_0), \qquad t - t_0 = 2aqB,$$

where $q$ is any integer of $F$. Eliminating $t$ from these equations, we have

$$y = 4a^2q^2B + 4aqt_0 + y_0.$$

The coefficient of $q^2$ is not zero, and hence this formula gives an infinitude of integral values of $y$. As in Case 1, we have $y \equiv y_0$ and $t \equiv t_0 \pmod{2a}$, so that the values of $y$ give integral values of $x$ in (16).

*Case* 3. $B^2 - 4AC = 0$; neither $A$ nor $C$ is negative when $F$ is a real field. In other words, we are now treating the case where the left side of equation (2) factors into two linear expressions, both being real when $F$ is real. Equation (16) can be written in the form

$$2ax + by + d = \pm(A^{1/2}y + C^{1/2}).$$

If both $A^{1/2}$ and $C^{1/2}$ are in $F$, then these linear equations have integral coefficients from $F$. As was remarked at the beginning of §1, one integral solution implies an infinite number.

If $A^{1/2}$ is in $F$, but $C^{1/2}$ is not, then obviously there is no integral solution in $F$, for such a solution would enable us to write $C^{1/2}$ as an element of $F$.

If $C^{1/2}$ is in $F$, but $A^{1/2}$ is not, then any solution $(x_0, y_0)$ must have $y_0=0$. Also, since $B=2A^{1/2}C^{1/2}$, and $B$ is in $F$, it follows that $C=0$. Hence the only possible solution is $y_0=0$, $x_0=-d/2a$.

If neither $A^{1/2}$ nor $C^{1/2}$ is in $F$, any integral solution $(x_0, y_0)$ must be such that $A^{1/2}y_0+C^{1/2}=0$, which fixes the value of $y_0$; and $x_0$ must therefore satisfy $2ax+by_0+d=0$, so that there cannot be more than one solution.

5. **The rational case.** We now prove Theorem 1; the coefficients of (2) are taken to be rational integers. The case in which (2) represents a pair of straight lines was treated in Case 3 in the last section. If (2) represents a parabola, then $A=0$ and $B\neq 0$. This was discussed in Case 2 in the last section. Hence we can complete the proof of Theorem 1 by treating the hyperbola. We prove this lemma.

LEMMA 5. *Let* (2) *represent an hyperbola, so that* $A>0$ *and* $B^2-4AC\neq 0$. *Then the asymptotes are essentially rational if and only if* $A$ *is a perfect square.*

First, if the asymptotes are rational, we can write (2) in the form

$$a(x+\alpha_1 y+\beta_1)(x+\alpha_2 y+\beta_2)=\delta, \tag{23}$$

where $\alpha_1$, $\alpha_2$, $\beta_1$, and $\beta_2$ are rational, and the asymptotes are obtained by equating to zero the expressions in parentheses. Equating coefficients in (2) and (23), we obtain

$$b=a(\alpha_1+\alpha_2), \qquad c=a\alpha_1\alpha_2.$$

Hence we can write

$$A=b^2-4ac=a^2(\alpha_1+\alpha_2)^2-4a^2\alpha_1\alpha_2=a_2(\alpha_1-\alpha_2)^2.$$

The integer $A$ is the square of a rational number, and consequently is the square of a rational integer.

Conversely, suppose that $A=k^2\neq 0$. In order to show that the asymptotes are rational, we exhibit them. Multiplying (2) by $4a$, we have

$$(2ax+by)^2-k^2y^2+4adx+4aey+4af=0.$$

Multiplying by $k^2$, and completing the squares, we obtain

$$(2akx+bky+dk)^2-(k^2y-2ae+bd)^2=T, \tag{24}$$

where $T$ is given by

$$d^2k^2-4afk^2-(2ae-bd)^2. \tag{25}$$

The asymptotes of the hyperbola are obtained by factoring the difference of two squares on the left of (24), and equating the factors to zero. It is obvious that they are rational lines, and this completes the proof of the lemma.

We now consider Case 1 of §4 in the light of Lemma 5, and see that we have proved that an hyperbola (2), with irrational asymptotes, has either no integral solutions or an infinite number. To complete the proof of Theorem 1, we must show that an hyperbola (2), with rational asymptotes, cannot have an infinite number of integral solutions.

Let $(x_0, y_0)$ be a point with integral coordinates lying on the hyperbola. Let equation (1), with rational integral coefficients, denote an asymptote. Then the distance from the point on the curve to the asymptote is

$$\left| \frac{lx_0 + my_0 + n}{(l^2 + m^2)^{1/2}} \right| \geqq \frac{1}{(l^2 + m^2)^{1/2}},$$

since $lx_0+my_0+n$ is a nonzero integer. But the asymptotes approach the curve, so that the points of the hyperbola whose distances from the adjacent asymptote are greater than any given positive quantity, must lie in a finite region of the plane. Consequently, only a finite number of points with integral coordinates lie on the hyperbola.

6. **The proof of Theorem 2.** Let the coefficients of (2) be integers of a real quadratic field. If (2) represents a point, then obviously it cannot have more than one integral solution. The situation in which (2) represents a pair of straight lines was treated in Case 3 of §4; a parabola in Case 2; an hyperbola, or an ellipse with $A$ not totally negative in Case 1. All that remains is the last statement of Theorem 2, concerning the ellipse with $A$ totally negative; we turn to this now.

Multiplying equation (2) by $4a$, we get

$$(2ax + by)^2 - Ay^2 + 4adx + 4aey + 4af = 0.$$

We multiply by $-A$, and complete the squares to arrive at

$$-AX^2 + Y^2 = (bd - 2ae)^2 - A(d^2 - 4af), \tag{26}$$

where

$$X = 2ax + by + d, \qquad Y = Ay + bd - 2ae. \tag{27}$$

Suppose that the quadratic field with which we are dealing is $R(m^{1/2})$, where $m$ is positive. Then the integer $A$, being totally negative, has the form

$$-p - qm^{1/2}, \qquad p > |qm^{1/2}| \geqq 0, \tag{28}$$

where $p$ and $q$ are rational integers, or perhaps the halves of odd rational integers in case $m \equiv 1 \pmod 4$. We are looking for integral values of $x$ and $y$ in $R(m^{1/2})$, so we suppose that $X = w + tm^{1/2}$, and $Y = u + vm^{1/2}$. Let the right side of equation (26) be $r + sm^{1/2}$. The quantities $w, t, u, v, r$, and $s$ are rational integers (or perhaps the halves of odd rational integers).

Substituting these values in (26), and equating the rational terms, we have the result

$$p(w^2 + t^2m) + 2qmwt + u^2 + v^2m = r. \tag{29}$$

The inequality in (28) enables us to write

$$p(w^2 + t^2m) \geqq 2 \mid qm^{1/2} \mid \cdot \mid wtm^{1/2} \mid = \mid 2qmwt \mid,$$

so that $r$ must not be negative if (29) is to have any solutions. Equation (29) implies that

$$u^2 + v^2m \leqq r, \qquad pw^2 + pmt^2 + 2qmwt \leqq r.$$

Clearly the first of these inequalities has only a finite number of solutions in integers (or halves of odd integers) $u$ and $v$. The same is true of the second inequality in $w$ and $t$, because the discriminant of the left side is

$$4q^2m^2 - 4mp^2 < 4q^2m^2 - 4m(mq^2) = 0,$$

by (28). Hence the number of integral solutions in $X$ and $Y$ of (26) is finite, and, by (27), the number of integral solutions of (2) is finite.

7. **Imaginary quadratic fields.** We now prove Theorem 3. The situation in which $B^2-4AC=0$, that is, in which the left side of (2) factors, was treated in Case 3 of §4. When $B^2-4AC\neq 0$, Cases 1 and 2 handle the situations with $A$ not a perfect square, and $A$ zero, respectively. All that remains to be proved, therefore, is that (2) cannot have an infinite number of solutions when $B^2-4AC\neq 0$ and $A$ is a perfect square in $F$, say $k^2$. We can proceed as in §5, and obtain equations (24) and (25); $T$ must be different from zero, since otherwise the left side of (2) would be factorable into two linear factors, contrary to hypothesis. We use the substitution

$$X = 2akx + bky + dk, \qquad Y = k^2y - 2ae + bd,$$

to write (24) in the form $X^2 - Y^2 = T$, from which we get

$$\mid X - Y \mid \cdot \mid X + Y \mid = \mid T \mid. \tag{30}$$

As in the last section, we show that there is only a finite number of solutions in $X$ and $Y$, and this implies the result we want. Now the positive rational integer $|T|$ can be factored into a pair of positive rational integers in but a finite number of ways. Any integral solution of (30) must correspond to one of these factorings. For any such factoring, say $|T| = rs$, we can write $|X-Y| = r$ and $|X+Y| = s$, or vice versa. But there is only a finite number of integers of any imaginary quadratic field with absolute value equal to a given rational integer. Hence we have only a finite number of pairs $(X-Y, X+Y)$ satisfying (30), and each pair gives at most one integral solution $(X, Y)$.

University of Illinois,
Urbana, Ill.

# ON THE PARTIAL SUMS OF HARMONIC DEVELOPMENTS AND OF POWER SERIES

BY

OTTO SZÁSZ

1. **Introduction.** Consider the class $E$ of power series $f(z)=\sum_0^\infty c_\nu z^\nu$, convergent for $|z|<1$ and such that $|f(z)|\leqq 1$. The following result is due to I. Schur and G. Szegö [5](1).

For any series of the class $E$,

$$|s_n(z)| \equiv \left|\sum_0^n c_\nu z^\nu\right| \leqq 1$$

in $|z|\leqq r_n$, but not always in $|z|<r_n+\epsilon$, $\epsilon>0$, where $r_n$ is the largest $r$ for which

$$T_n(r,\theta) = \frac{1}{2} + \sum_1^n r^\nu \cos \nu\theta \geqq 0 \qquad \text{for all } \theta.$$

The $r_n$ are non-decreasing,

$$r_n > 1 - \frac{\log 2n}{n}, \qquad n = 1, 2, 3, \cdots,$$

$$r_n = 1 - \frac{\log 2n - \log\log 2n + \epsilon_n}{n}, \qquad \lim_{n\to\infty} \epsilon_n = 0.$$

We obtain the same constant $r_n$ if we assume $Rf(z)\geqq 0$ and require $Rs_n(z)\geqq 0$. Here $Ru$ means the real part of $u$; $Iu$ will denote the imaginary part.

In what follows, we consider harmonic sine developments

$$H(r,\theta) = \sum_1^\infty b_\nu r^\nu \sin \nu\theta,$$

convergent for $0<r<1$, and non-negative for $0<\theta<\pi$. Evidently there exists an $R_n$ with the following properties:

(a) Whenever

$$H(r,\theta) \geqq 0, \qquad 0<r<1;\ 0<\theta<\pi, \tag{1.1}$$

then,

$$s_n(r,\theta) \equiv \sum_1^n b_\nu r^\nu \sin\nu\theta \geqq 0, \quad 0<r\leqq R_n;\ 0<\theta<\pi. \tag{1.2}$$

---

Presented to the Society, January 1, 1941; received by the editors March 10, 1941, and in revised form, May 7, 1941. The author is indebted to the referee for valuable suggestions.

(1) Numbers in brackets refer to the literature at the end of this paper.

(b) For any $\epsilon>0$ we can find an $H$ satisfying (1.1) and such that $s_n(r, \theta)$ becomes negative for some $\theta$ and some $r<R_n+\epsilon$.

We denote the class of harmonic functions satisfying (1.1) by $T$. On writing $f(z)=\sum_1^\infty b_\nu z^\nu$, the power series $f(z)$ is regular in $|z|<1$, has all its coefficients real, and $If(z)\geqq 0$ in $|z|<1$, $Iz>0$. The class $T$ has been discussed by Rogosinski [4]; the function $f(z)$ is called typically real. (Cf. also S. Mandelbrojt [2].)

One of the results of the present paper is

$$R_n = 1 - \frac{3\log n}{n} + \frac{\log\log n}{n} + O(1/n), \qquad \text{as } n \to \infty.$$

M. S. Robertson [3] gave the erroneous estimate

$$R_n \geqq 1 - 2\log n/n \qquad \text{for } n > 12.$$

His calculation yields however, as is seen easily, $R_n \geqq 1-4\log n/n$, for $n>n_0$(2). We then apply the properties of $R_n$ to Fourier series of convex functions and to a certain class of power series.

Note that if $\phi(\theta)\sim\sum_1^\infty b_\nu \sin \nu\theta$, $\phi(\theta)\geqq 0$, $0<\theta<\pi$, then

$$\begin{aligned}
H(r, \theta) &= \frac{2}{\pi}\int_0^\pi \phi(x)\left(\sum_1^\infty r^\nu \sin \nu\theta \sin \nu x\right)dx \\
&= \frac{1}{\pi}\int_0^\pi \phi(x)\sum_1^\infty r^\nu[\cos \nu(\theta - x) - \cos \nu(\theta + x)]dx \\
&= \frac{1-r^2}{2\pi}\int_0^\pi \phi(x)\left(\frac{1}{1-2r\cos(\theta-x)+r^2} - \frac{1}{1-2r\cos(\theta+x)+r^2}\right)dx \\
&= \frac{2r(1-r^2)}{\pi}\sin\theta\int_0^\pi \frac{\phi(x)\sin x\,dx}{[1-2r\cos(\theta-x)+r^2][1-2r\cos(\theta+x)+r^2]}.
\end{aligned}$$

Hence $H(r, \theta)$ belongs to the class $T$. (Cf. Zygmund [8, p. 57].)

2. **Characterization of $R_n$.** We quote the following lemma, due to Fejér (Turán [7]).

LEMMA 1. *In order that*

$$\sum_{\nu=1}^n \lambda_\nu \sin \nu x \sin \nu y \geqq 0 \quad \textit{for } 0 < x < \pi,\ 0 < y < \pi,$$

*it is necessary and sufficient that*

$$\sum_1^n \nu\lambda_\nu \sin \nu\theta \geqq 0 \qquad \textit{for } 0 < \theta < \pi.$$

(2) Robertson, Annals of Mathematics, (2), vol. 42 (1941), pp. 829–838.

We now prove

THEOREM 1. *The quantity $R_n$ as defined in* §1 *is the largest $r$ for which*

$$S_n(r, \theta) \equiv \sum_1^n \nu r^\nu \sin \nu\theta \geqq 0 \qquad \textit{for} \quad 0 < \theta < \pi. \tag{2.1}$$

We have for $0<\rho<1$,

$$\rho^\nu b_\nu = \frac{2}{\pi}\int_0^\pi H(\rho, x) \sin \nu x \, dx, \qquad \nu = 1, 2, 3, \cdots ;$$

hence

$$s_n(r, \theta) = \frac{2}{\pi}\int_0^\pi H(\rho, x)\left(\sum_1^n \left(\frac{r}{\rho}\right)^\nu \sin \nu\theta \sin \nu x\right) dx.$$

For any $r<R_n$ we can choose $\rho<1$ so that $r/\rho<R_n$; we then obtain by Lemma 1 (for $\lambda_\nu=r^\nu$) $s_n(r, \theta)\geqq 0$ for any $r<R_n$ and for $0<\theta<\pi$; hence (a) holds for $r\leqq R_n$. Conversely, for the function

$$H(r, \theta) = \sum_1^\infty \nu r^\nu \sin \nu\theta = r \sin \theta \frac{1 - r^2}{(1 - 2r \cos \theta + r^2)^2} > 0$$

the function

$$s_n(r, \theta) = \sum_1^n \nu r^\nu \sin \nu\theta$$

becomes negative for any $r>R_n$ and for some $\theta$ in $(0, \pi)$. This proves Theorem 1. To estimate $R_n$ we first give another characterization for it. An easy calculation yields

$$\begin{aligned}
&\frac{(1 - 2r \cos \theta + r^2)^2}{r \sin \theta} \sum_1^n \nu r^\nu \sin \nu\theta \\
&\qquad = 1 - r^2 - (n + 1)r^{n+2} \cdot \frac{\sin (n - 1)\theta}{\sin \theta} + r^{n+1}(2n + 2 + nr^2)\frac{\sin n\theta}{\sin \theta} \\
&\qquad\qquad - r^n(n + 1 + 2nr^2)\frac{\sin (n + 1)\theta}{\sin \theta} + nr^{n+1}\frac{\sin (n + 2)\theta}{\sin \theta} \\
&\qquad \equiv C_n(r, \theta).
\end{aligned}$$

This furnishes

THEOREM 2. *$R_n$ is the largest $r$ for which*

$$C_n(r, \theta) \geqq 0 \qquad \textit{for all } \theta.$$

Evidently

$$C_n(r, \pi) = 1 - r^2 + (n^2 - 1)r^{n+2}(-1)^{n-1} + nr^{n+1}(2n + 2 + nr^2)(-1)^{n-1}$$
$$+ (n+1)r^n(n + 1 + 2nr^2)(-1)^{n-1} + n(n+2)r^{n+1}(-1)^{n-1}$$
$$= 1 - r^2 + (-1)^{n-1}\{(n^2 - 1)r^{n+2} + nr^{n+1}(2n + 2 + nr^2)$$
$$+ (n+1)r^n(n + 1 + 2nr^2) + n(n+2)r^{n+1}\}.$$

Thus

$$C_n(r, \theta) \geqq 1 - r^2 - \{(n^2 - 1)r^{n+2} + nr^{n+1}(2n + 2 + nr^2)$$
$$+ (n+1)r^n(n + 1 + 2nr^2) + n(n+2)r^{n+1}\}, \quad n \geqq 1,$$

and equality holds if $n = 2k$, and $\theta = \pi$. This yields

THEOREM 3. *Denote the unique positive root of the equation*

$$p_n(r) \equiv 1 - r^2 - (n+1)^2 r^n - n(3n+4)r^{n+1} - (3n^2 + 2n - 1)r^{n+2} - n^2 r^{n+3} = 0$$

*by $\rho_n$. Then $R_n \geqq \rho_n$, and equality holds for $n = 2k$, $k \geqq 1$.*

Note that $p_n(0) = 1$, $p_n(1) < 0$, $p_n'(r) < 0$. Hence $\rho_n$ is unique and

$$0 < \rho_n < 1. \tag{2.2}$$

Evidently $p_n(-1) = 0$, hence $1 + r$ can be factored out, and we get

$$\frac{p_n(r)}{1+r} = 1 - r - (n+1)^2 r^n - (2n^2 + 2n - 1)r^{n+1} - n^2 r^{n+2} \equiv q_n(r), \tag{2.3}$$

so that $q_n(\rho_n) = 0$.

3. **Estimation of $\rho_n$ and $R_n$.** Direct calculation gives

$$R_1 = 1; \rho_1 = 0.182 \cdots.$$

Also $\rho_2 = R_2$, and

$$S_2(r, \theta) = r \sin\theta + 2r^2 \sin 2\theta = r \sin\theta(1 + 4r\cos\theta),$$

which yields by Theorem 1: $R_2 = 1/4 = \rho_2$. A similar calculation yields $R_3 = 2^{1/2}/3$.

We shall prove

$$\rho_n = 1 - \frac{3\log n}{n} + \frac{\log\log n + \log 3/4 + \epsilon_n}{n}, \qquad \epsilon_n \to 0 \text{ as } n \to \infty. \tag{3.1}$$

Let $c$ be a constant, and

$$r_n(c) = 1 - \frac{3\log n}{n} + \frac{\log\log n + c}{n}; \tag{3.2}$$

then from

$$\log(1 - x) = -x + O(x^2) \qquad \text{as} \quad x \to 0,$$

we conclude

$$\begin{aligned}(3.3)\qquad \{r_n(c)\}^n &= \exp\{-3\log n + \log\log n + c + O(n^{-1}\log^2 n)\} \\ &= n^{-3}\log n\cdot e^c\{1 + O(n^{-1}\log^2 n)\} \quad \text{as} \quad n \to \infty.\end{aligned}$$

Furthermore, from (2.3), (3.2), and (3.3)

$$q_n\{r_n(c)\} = \frac{3\log n}{n} - \frac{\log\log n + c}{n} - \frac{4\log n}{n}\cdot e^c\{1 + o(1)\},$$

hence

$$\frac{nq_n\{r_n(c)\}}{\log n} \to 3 - 4e^c \qquad \text{as } n \to \infty.$$

Thus for

$$c = \log 3/4 + \epsilon,$$

$\epsilon$ a given small number, and for sufficiently large values of $n$

$$\operatorname{sgn} q_n\{r_n(c)\} = \operatorname{sgn} \epsilon,$$

from which follows (3.1).

We have thus proved

THEOREM 4. *If* $\rho_n > 0$ *and* $p_n(\rho_n) = 0$, *then*

$$\rho_n = 1 - \frac{3\log n}{n} + \frac{\log\log n + \log 3/4 + \epsilon_n}{n}, \qquad \textit{where } \epsilon_n \to 0 \textit{ as } n \to \infty.$$

4. **Derivation of an asymptotic estimate for $R_n$.** On writing

$$R_n = 1 - \frac{3\log n}{n} + \frac{\log\log n + \delta_n}{n}, \qquad n > 1,$$

it follows from Theorem 3 that

$$\delta_n \geqq \log 3/4 + \epsilon_n,$$

and equality holds for $n = 2k$, $k \geqq 1$; hence from Theorem 4

$$\liminf_{n\to\infty} \delta_n = \log 3/4, \qquad \lim_{k\to\infty} \delta_{2k} = \log 3/4.$$

It remains to give an estimate for $R_{2k-1}$ from above.

If for a particular value of $\theta$ and $r$, $C_{2k-1}(r, \theta) < 0$, then by Theorem 2, evidently $R_{2k-1} < r$. We now choose $\theta = \pi - (3\pi/4k)$; then

$$
\begin{aligned}
C_{2k-1}(r,\theta) &= 1 - r^2 + \frac{1}{\sin(3\pi/4k)}\left\{2kr^{2k+1}\sin\frac{3(k-1)\pi}{2k}\right. \\
&\quad + r^{2k}[4k + (2k-1)r^2]\sin\frac{3(2k-1)\pi}{4k} \\
&\quad + r^{2k-1}[2k + 2(2k-1)r^2]\sin\frac{3\pi}{2} \\
&\quad \left. + (2k-1)r^{2k}\sin\frac{3(2k+1)\pi}{4k}\right\} \\
&= 1 - r^2 - \frac{1}{\sin(3\pi/4k)}\left\{2kr^{2k+1}\cos\frac{3\pi}{2k}\right. \\
&\quad + r^{2k}[4k + (2k-1)r^2]\left[\cos\frac{3\pi}{4k}\right] \\
&\quad \left. + 2kr^{2k-1} + (4k-2)r^{2k+1} + (2k-1)r^{2k}\cos\frac{3\pi}{4k}\right\} \\
&< 1 - r^2 - \frac{4k}{3\pi}\left\{2kr^{2k+1}\left(1 - \frac{9\pi^2}{8k^2}\right)\right. \\
&\quad + r^{2k}[4k + (2k-1)r^2]\left(1 - \frac{9\pi^2}{32k^2}\right) \\
&\quad \left. + 2kr^{2k-1} + (4k-2)r^{2k+1} + (2k-1)r^{2k}\left(1 - \frac{9\pi^2}{32k^2}\right)\right\}, \qquad k \geqq 3
\end{aligned}
$$

(since $\cos x > 1 - x^2/2$ for all $x$). Hence

$$
\begin{aligned}
C_{2k-1}(r,\theta) &< 1 - r^2 - (2k/5)\{kr^{2k+1} + 2kr^{2k} + (k-1/2)r^{2k+2} + 2kr^{2k-1} \\
&\quad + (4k-2)r^{2k+1} + (k-1/2)r^{2k}\}, \qquad k \geqq 5,
\end{aligned}
$$

thus

$$C_{2k-1}(r,\theta) < 1 - r^2 - (2k/5)(11k-3)r^{2k+2} < 1 - r^2 - 4k^2r^{2k+2}.$$

Choosing $r$ so that

$$1 - r^2 - 4k^2r^{2k+2} \leqq 0, \tag{4.1}$$

we get

$$C_{2k-1}(r,\theta) < 0, \qquad R_{2k-1} < r.$$

To find an upper bound for $r$, we put

$$r = 1 - \frac{3\log(2k-1)}{2k-1} + \frac{\log\log(2k-1) + c}{2k-1}; \tag{4.2}$$

we obtain as in (3.3)

$$\begin{aligned} r^{2k-1} &= \exp\left\{-3\log(2k-1)+\log\log(2k-1)+c+O(k^{-1}\log^2 k)\right\} \\ &= (2k-1)^{-3}\log(2k-1)\cdot e^c\left\{1+O(k^{-1}\log^2 k)\right\}. \end{aligned}$$

Thus, using (4.2),

$$\begin{aligned} \frac{4k^2r^{2k+2}}{1-r^2} &= \left(\frac{2k}{2k-1}\right)^2\cdot\frac{r^3}{1+r}\cdot\frac{r^{2k-1}(2k-1)^2}{1-r} \\ &= \frac{1+o(1)}{2+o(1)}\cdot\frac{1}{3}e^c\left\{1+o(1)\right\}\to\frac{1}{6}e^c \qquad \text{as } k\to\infty. \end{aligned}$$

Hence (4.1) is satisfied for all sufficiently large $k$ provided $e^c/6>1$, that is, $c>\log 6$. It now follows that $\limsup_{k\to\infty}\delta_{2k-1}\leqq 6$. Summarizing we have

THEOREM 5. *Let*

$$R_n = 1-\frac{3\log n}{n}+\frac{\log\log n+\delta_n}{n}, \qquad n>1;$$

*then* $\lim_{k\to\infty}\delta_{2k}=\log 3/4$, *and*

$$\log 3/4 \leqq \liminf_{k\to\infty}\delta_{2k-1}\leqq\limsup_{k\to\infty}\delta_{2k-1}\leqq 6.$$

5. **Application to Fourier series.** Consider the roof-function

$$\frac{2b}{a(\pi-a)}\sum_1^\infty\frac{\sin\nu a\sin\nu\theta}{\nu^2}=\begin{cases}\dfrac{b\theta}{a} & \text{for } 0\leqq\theta\leqq a,\\[2ex] b\,\dfrac{\pi-\theta}{\pi-a} & \text{for } a\leqq\theta\leqq\pi,\end{cases}$$

where $0<a<\pi$, $0<b$, and the corresponding harmonic function

$$\frac{2b}{a(\pi-a)}\sum_1^\infty r^\nu\frac{\sin\nu a\sin\nu\theta}{\nu^2}=H(r;a,b).$$

Denote its partial sums by

$$H_n(r,\theta)=\frac{2b}{a(\pi-a)}\sum_1^n r^\nu\frac{\sin\nu a\sin\nu\theta}{\nu^2};$$

then

$$\frac{\partial^2H_n(r,\theta)}{\partial\theta^2}=-\frac{2b}{a(\pi-a)}\sum_1^n r^\nu\sin\nu a\sin\nu\theta\leqq 0$$

for $0<r\leqq R_n$, $0<\theta<\pi$, by Lemma 1 and Theorem 1. Hence $H_n(r,\theta)$ is con-

vex upwards for $0<\theta<\pi$, $r\leqq R_n$; but not convex for $r>R_n$. The same is true for the limiting cases $a\to 0$ and $a\to\pi$. In which cases

$$H(r;\,0,\,b)=\frac{2b}{\pi}\sum_1^\infty r^\nu\frac{\sin\nu\theta}{\nu},$$

$$H(r;\,\pi,\,b)=\frac{2b}{\pi}\sum_1^\infty r^\nu\frac{\sin\nu(\pi-\theta)}{\nu}.$$

Moreover every polygon convex upwards and lying above the axis of abscissae is expressible as a finite sum with positive coefficients of roof-functions. Hence the partial sums of the corresponding harmonic development are convex upwards for $r\leqq R_n$. Finally any function positive in $0<\theta<\pi$, and convex upwards can be approximated uniformly by such polygons; hence we have

THEOREM 6. *If* $f(\theta)>0$ *in* $0<\theta<\pi$, *and is convex upwards, and if* $f(\theta)\sim\sum_1^\infty b_\nu \sin\nu\theta$, *then* $\sum_1^n r^\nu b_\nu \sin\nu\theta$ *is convex upwards in* $0<\theta<\pi$, $r\leqq R_n$; *but not always for* $r<R_n+\epsilon$, $\epsilon>0$.

6. **Cosine series.** We now consider the cosine series of the step function

$$\frac{2b}{\pi}\left\{\frac{\pi-a}{2}-\sum_1^\infty\frac{\sin\nu a\cos\nu\theta}{\nu}\right\}=\begin{cases}0 & \text{for } 0\leqq\theta<a,\\ b & \text{for } a<\theta\leqq\pi,\end{cases}$$

where $0<a<\pi$, $b>0$; and the corresponding harmonic development

$$K(r,\theta)=\frac{b}{\pi}(\pi-a)-\frac{2b}{\pi}\sum_1^\infty r^\nu\frac{\sin\nu a\cos\nu\theta}{\nu}.$$

For the partial sums $K_n(r,\theta)$ of this series we have

$$\frac{\partial K_n(r,\theta)}{\partial\theta}=\frac{2b}{\pi}\sum_1^n r^\nu\sin\nu a\sin\nu\theta\geqq 0\qquad\text{for } 0<r\leqq R_n,\ 0<\theta<\pi;$$

hence $K_n(r,\theta)$ is monotonic increasing in the same domain; $R_n$ cannot be replaced by $R_n+\epsilon$, $\epsilon>0$. The same statement for any monotonic increasing function follows now in an obvious way. Hence we have

THEOREM 7. *If* $f(\theta)$ *is monotonic in* $0<\theta<\pi$, *and*

$$f(\theta)\sim a_0/2+\sum_1^\infty a_\nu\cos\nu\theta,$$

*then the nth partial sum of* $a_0/2+\sum_1^\infty a_\nu r^\nu\cos\nu\theta$ *is monotonic in the same sense for* $0<r\leqq R_n$, *and here* $R_n$ *cannot be replaced by* $R_n+\epsilon$, $\epsilon>0$.

7. **Curves convex in direction of the $v$-axis.** We say that a curve in the $(u,v)$-plane is convex in the direction of the $v$-axis if any parallel to the $v$-axis

has at most two points in common with the curve. This class of mappings was considered by L. Fejér [1] and the author [6]. We now prove

THEOREM 8. *Suppose the power series* $\sum_0^\infty a_\nu z^\nu = f(z) = w = u + iv$ *is regular in* $|z| < 1$, *and all* $a_\nu$ *are real. Suppose further that the images* $K_r$ *of the circles* $|z| = r$, $0 < r < 1$, *are convex in the direction of the v-axis (thus* $f(z)$ *is univalent). Then the partial sum* $\sum_0^n a_\nu z^\nu$ *has the same property in* $|z| \leqq R_n$, *but—in general—not in a larger circle.*

For the proof we may assume without loss of generality that the upper half of the circle $|z| < 1$ is mapped onto the upper half of the image in the $w$-plane. On writing $w(e^{i\theta}) = u(\theta) + iv(\theta) \sim \sum_0^\infty a_\nu \cos \nu\theta + i\sum_0^\infty a_\nu \sin \nu\theta$, we find that $v(\theta)$ is positive for $0 < \theta < \pi$, and (from the assumption) $u(\theta)$ is decreasing in the same interval. Our theorem follows now from Theorems 5 and 7.

8. **Conclusion.** Suppose $f(z) = \sum_1^\infty b_\nu z^\nu$ is a typically real function, that is,

$$\sum_1^\infty b_\nu r^\nu \sin \nu\theta \geqq 0 \qquad \text{for} \quad 0 < r < 1,\ 0 < \theta < \pi.$$

Then the Riesz means of second order

$$P_n(z) = (n+1)^{-2} \sum_{\nu=1}^{n} (n - \nu + 1)^2 b_\nu z^\nu, \qquad n \geqq 1,$$

are typically real in $|z| \leqq 1$ (Szász [6]; cf. Theorem 1). Evidently $\lim_{n\to\infty} P_n(z) = f(z)$ in $|z| < 1$, uniformly in $|z| \leqq r$, $r < 1$. Another such sequence of polynomials is

$$s_n(R_n z) = \sum_{\nu=1}^{n} b_\nu R_n^\nu z^\nu, \qquad n \geqq 1.$$

These polynomials are typical real in $|z| \leqq 1$ by property (a) of §1. Furthermore for $|z| \leqq r < 1$

$$\begin{aligned} | f(z) - s_n(R_n z) | &\leqq \sum_1^n | b_\nu | r^\nu (1 - R_n^\nu) + \sum_{n+1}^\infty | b_\nu | r^\nu \\ &< (1 - R_n) \sum_1^n \nu | b_\nu | r^\nu + \sum_{n+1}^\infty | b_\nu | r^\nu \to 0, \text{ as } n \to \infty. \end{aligned}$$

Hence

$$\lim_{n\to\infty} s_n(R_n z) = f(z)$$

uniformly in $|z| \leqq r < 1$.

## REFERENCES

1. L. Fejér, *Neue Eigenschaften der Mittelwerte bei den Fourierreihen*, Journal of the London Mathematical Society, vol. 8 (1933), pp. 53–62.

2. S. Mandelbrojt, *Quelques remarques sur les fonctions univalentes*, Bulletin des Sciences Mathématiques, (2), vol. 58 (first part) (1934), pp. 185–200.

3. M. S. Robertson, *On the theory of univalent functions*, Annals of Mathematics, (2), vol. 37 (1936), pp. 374–408.

4. W. Rogosinski, *Über positive harmonische Entwicklungen und typischreelle Potenzreihen*, Mathematische Zeitschrift, vol. 35 (1932), pp. 93–121.

5. I. Schur and G. Szegö, *Über die Abschnitte einer im Einheitskreise beschränkten Potenzreihe*, Sitzungsberichte der Preussischen Akademie, 1923, pp. 545–560.

6. O. Szász, *On the Cesàro and the Riesz means of Fourier series*, Compositio Mathematica, vol. 7 (1939), pp. 112–122.

7. P. Turán, *Über die montone Konvergenz der Cesàro-Mittel bei Fourier- und Potenzreihen*, Proceedings of the Cambridge Philosophical Society, vol. 34, Part II (1938), pp. 134–143.

8. A. Zygmund, *Trigonometrical Series*, 1935.

UNIVERSITY OF CINCINNATI,
CINCINNATI, OHIO

# HYPERSPACES OF A CONTINUUM

BY

J. L. KELLEY

**Introduction.** Among the topological invariants of a space $X$ certain spaces have frequently been found valuable. The space of all continuous functions on $X$ and the space of mappings of $X$ into a circle are noteworthy examples. It is the purpose of this paper to study two particular invariant spaces associated with a compact metric continuum $X$; namely, $2^X$, which consists of all closed nonvacuous subsets of $X$, and $\mathcal{C}(X)$, which consists of closed connected nonvacuous subsets(1). The aim of this study is twofold. First, we wish to investigate at length the topological properties of the hyperspaces, and, second, to make use of their structure to prove several general theorems.

If $X$ is a compact metric continuum it is known that: $2^X$ is Peanian if $X$ is Peanian [7], and conversely [8]; $2^X$ is always arcwise connected [1]; $2^X$ is the continuous image of the Cantor star [4]; if $X$ is Peanian, each of $2^X$ and $\mathcal{C}(X)$ is contractible in itself [9]; and if $X$ is Peanian, $2^X$ and $\mathcal{C}(X)$ are absolute retracts [10].

In §§1–5 of this paper further topological properties are obtained. In particular: $2^X$ has vanishing homology groups of dimension greater than 0, both hyperspaces have very strong higher local connectivity and connectivity properties—including local $p$-connectedness in the sense of Lefschetz for $p>0$, and, the question of dimension is resolved except for the dimension of $\mathcal{C}(X)$ when $X$ is non-Peanian. All of the results of the preceding paragraph for $2^X$ are shown simultaneously for $2^X$ and $\mathcal{C}(X)$ in the course of the development.

In §6 a characterization of local separating points in terms of $\mathcal{C}(X)$ is obtained and a theorem of G. T. Whyburn deduced. In §7 it is shown that for a continuous transformation $f(X)=Y$ we may under certain conditions find $X_0 \subset X$, with $X_0$ closed and of dimension 0, such that $f(X_0)=Y$. In §8 this result is utilized in the study of Knaster continua. In order that $X$ be a Knaster continuum it is necessary and sufficient that $\mathcal{C}(X)$ contain a unique arc between every pair of elements. If there exist Knaster continua of dimension greater than 1 then there exist infinite-dimensional Knaster continua.

---

Presented to the Society in three parts, the first under the title *On the hyperspaces of a given space* on December 28, 1939; the second under the present title on April 26, 1940; the third under the title *A theorem on transformations* on December 30, 1940; received by the editors May 15, 1941. The major part of the material in this paper was contained in the author's dissertation, University of Virginia, June 1940.

(1) For topologization of these spaces and for definitions of terms used in the introduction see the text. A bibliography is given at the end of the article. Numbers in square brackets refer to the bibliography.

The author wishes to express his gratitude to Professor G. T. Whyburn for his help and encouragement in the preparation of this paper.

1. **Preliminaries.** Throughout the following, $X$ will denote a compact metric continuum. The letters $a$, $b$, $c$ will stand for elements of $X$. For $a, b \in X$, $\rho(a, b)$ is the distance from $a$ to $b$. Given a collection $a_i$, $a_i \in X$, $\{a_i\}$ denotes the subset of $X$ whose elements are the $a_i$. In particular, $\{a\}$ is the subset of $X$ consisting of the one element $a$.

The letters $A$, $B$, $C$ stand for closed subsets of $X$. By $2^X$ we mean the space of all closed, nonvacuous subsets of $X$ metricized by the Hausdorff metric (that is, $\rho^1(A, B) = \text{g.l.b.}\ \{\epsilon\}$ for all $\epsilon$ such that $A \subset V_\epsilon(B)$ and $B \subset V_\epsilon(A)$, where $V_\epsilon(A)$ is the sum of all open $\epsilon$-spheres about points of $A$ [2]). If $A \in 2^X$ then $A \subset X$. The closed subspace of $2^X$ consisting of subcontinua of $X$ is $\mathcal{C}(X)$.

Similarly $2^{2^X}$ consists of closed, nonvacuous subsets $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ of $2^X$, with Hausdorff distance $\rho^2$. If $\mathcal{A} \in 2^{2^X}$ then $\mathcal{A} \subset 2^X$.

For $A \in 2^X$ we define $\phi(A) = \{\{a_i\}\}$, $a_i \in A$. That is, $\phi(A)$ is the subset of $2^X$ consisting of all elements $\{a\}$ of $2^X$ where $a \in A$. In particular $\phi(X)$ is the set of all sets $\{a\}$. We always have $\phi(A) \subset 2^X$ and $\phi(A) \in 2^{2^X}$. For any $A \in 2^X$, $\phi(A)$ is isometric with $A$. Similarly, for $\mathcal{A} \in 2^{2^X}$, $\phi(\mathcal{A})$ denotes the subset of $2^{2^X}$ consisting of elements $\{A\}$, $A \in \mathcal{A}$. We have $\phi(\mathcal{A}) \subset 2^{2^X}$ and in particular $\phi(2^X) \subset 2^{2^X}$.

For $\mathcal{A} \subset 2^X$ we define $\sigma(\mathcal{A}) = \sum A$ for all $A \in \mathcal{A}$. For every $\mathcal{A}$, $\sigma(\mathcal{A}) \subset X$. Actually $\sigma$ is a continuous mapping of $2^{2^X}$ onto $2^X$. Further:

1.1. Lemma. (a) *$\sigma$ is a contraction*, (b) *$\phi\sigma$ is a retraction of $2^{2^X}$ onto $\phi(2^X)$.*

**Proof.** First, for $\mathcal{A} \in 2^{2^X}$, $\sigma(\mathcal{A})$ is closed. Suppose $a_i \in \sigma(\mathcal{A})$, and $\lim a_i = a$. Choose $A_i$, $a_i \in A_i \in \mathcal{A}$. We can suppose $\lim A_i = A$. Since $\mathcal{A}$ is closed, $A \in \mathcal{A}$ and $a \in A \in \mathcal{A}$. Hence $a \in \sigma(\mathcal{A})$.

Second, suppose $\rho^1(\sigma(\mathcal{A}), \sigma(\mathcal{B})) = d$. We can choose in one of $\sigma(\mathcal{A})$, $\sigma(\mathcal{B})$, say in $\sigma(\mathcal{A})$, a point $a$ which is at least $d$ distance from every point of $\sigma(\mathcal{B})$. Choose $A$, $a \in A \in \mathcal{A}$. This set $A$ is then at least $d$ Hausdorff distance from every set $B \in \mathcal{B}$. Hence $\rho^2(\mathcal{A}, \mathcal{B}) \geqq d$ and $\sigma$ is shown to be a contraction. That $\sigma$ followed by $\phi$ leaves every element of $\phi(2^X)$ fixed is clear.

1.2. Lemma. *If $\mathcal{A}$ is a subcontinuum of $2^X$ and $\mathcal{A} \cdot \mathcal{C}(X) \neq 0$ then $\sigma(\mathcal{A})$ is a continuum.*

**Proof.** Choose $A \in \mathcal{A} \cdot \mathcal{C}(X)$. Suppose $\sigma(\mathcal{A}) = A_1 + A_2$ is a separation, with $A \subset A_1$. Then both the subset $\mathcal{A}_1$ of $\mathcal{A}$ consisting of all elements contained in $A_1$ and the subset $A_2$ of all elements intersecting $A_2$ are closed and nonvacuous. But $\mathcal{A}_1 + \mathcal{A}_2 = \mathcal{A}$, a continuum, and $A_1 \cdot A_2 = 0$. We then have a contradiction.

It is possible to define[2] a real-valued function $\mu(A)$, continuous on $2^X$,

[2] See H. Whitney, *Regular families of curves*, Annals of Mathematics, (2), vol. 34 (1933), p. 246.

with the properties:

1.3. *If* $A \subset B$, $A \neq B$ *then* $\mu(A) < \mu(B)$.

1.4. $\mu(X) = 1$, *and for any* $a \in X$, $\mu(\{a\}) = 0$.

For convenience, we shall suppose throughout that $\mu(A)$ is a certain fixed function with these properties. Since $2^X$ is compact we can further state:

1.5. LEMMA. *There exists* $\eta(\epsilon) > 0$ *such that if* $A$, $B \in 2^X$, $A \subset B$ *and* $\mu(B) - \mu(A) < \eta(\epsilon)$ *then* $\rho^1(A, B) < \epsilon$.

2. **Segments in** $2^X$. Let $A_0$, $A_1 \in 2^X$. A *segment* from $A_0$ to $A_1$ is a continuous mapping $A_t$ of the interval $[0, 1]$ into $2^X$ which satisfies the two conditions:

2.1. $\mu(A_t) = (1-t)\mu(A_0) + t\mu(A_1)$.

2.2. *If* $t' < t''$, *then* $A_{t'} \subset A_{t''}$.

2.3. LEMMA. *Given* $A_0$, $A_1 \in 2^X$, *there exists a segment from* $A_0$ *to* $A_1$ *if and only if* $A_0 \subset A_1$ *and every component of* $A_1$ *intersects* $A_0$.

**Proof.** First, suppose that $A_t$ is a segment from $A_0$ to $A_1$. If $A_1 = B_0 + B_1$ is a separation of $A_1$ such that $A_0 \subset B_0$, then the subset of $[0, 1]$ consisting of all $t$ such that $A_t \subset B_0$ and the subset defined by $A_t \cdot B_1 \neq 0$ are closed, disjoint and they cover $[0, 1]$. Hence $B_1 = 0$.

Second, suppose $A_0$, $A_1 \in 2^X$, $A_0 \subset A_1$ and every component of $A_1$ intersects $A_0$. Consider the collection of all sets $\mathcal{A} \subset 2^X$ which have the two properties:

2.4. *If* $B \in \mathcal{A}$ *then* $A_0 \subset B \subset A_1$ *and every component of* $B$ *intersects* $A_0$.

2.5. *If* $B_0$, $B_1 \in \mathcal{A}$ *then either* $B_0 \subset B_1$ *or* $B_0 \supset B_1$.

The sum of a monotone family of sets $\mathcal{A}$ of this collection is surely a member of the collection. Hence there must exist a member $\mathcal{A}_0$ which is saturated with respect to 2.4 and 2.5. Since the closure of $\mathcal{A}_0$ also satisfies 2.4 and 2.5, it follows that $\mathcal{A}_0$ is closed.

We now define for $t$, $0 \leqq t \leqq 1$, $A_t$ to be that element of $\mathcal{A}_0$ if it exists, such that $\mu(A_t) = (1-t)\mu(A_0) + t\mu(A_1)$. By 2.5 we see that $A_t$ is 1-1 and continuity follows from the continuity of the $\mu$ function. Now the proof will be complete if we show that $A_t$ is defined for every $t$, $0 \leqq t \leqq 1$, or—what is the same—that for $A_{t'}$, $A_{t''} \in \mathcal{A}_0$, $0 \leqq t' < t'' \leqq 1$, there exists $A \in \mathcal{A}_0$ such that $\mu(A_{t'}) < \mu(A) < \mu(A_{t''})$. Because of the maximal character of $\mathcal{A}_0$ it is sufficient to show that there exists some $A \in 2^X$ satisfying $A_{t'} \subset A \subset A_{t''}$, $\mu(A_{t'}) < \mu(A) < \mu(A_{t''})$ with every component of $A$ intersecting $A_{t'}$. Choose then $\epsilon > 0$ so that $\overline{V_\epsilon(A_{t'})}$ fails to contain $A_{t''}$, and let $A$ consist of the components of $A_{t''} \cdot \overline{V_\epsilon(A_{t'})}$ which intersect $A_{t'}$. Now some component of $A_{t''}$ is not contained in $\overline{V_\epsilon(A_{t'})}$ and hence $A_{t'}$ is a proper subset of $A$, while $A$ is surely a proper subset of $A_{t''}$. The required properties follow.

Since any subarc of a segment is, with proper parametrization, a segment, we have

2.6. LEMMA. *If $A \in \mathcal{C}(X)$ then every segment with $A$ as beginning is contained in $\mathcal{C}(X)$.*

The *Cantor star* is the plane set obtained by joining with a straight line every point of a discontinuum $D$ which lies on the $x$-axis to the point $(0, 1)$ on the $y$-axis. Each point of the star can be identified by a point $x \in D$ and a coordinate $y$, $0 \leqq y \leqq 1$.

The following theorem has been proved by Mazurkiewicz for $2^X$. (See [4] and also [1].)

2.7. THEOREM. *Each of $2^X$ and $\mathcal{C}(X)$ is the continuous image of the Cantor star, and hence arcwise connected*(3).

**Proof.** We first show that the set $\Sigma$ of all segments in $2^X$ and the set $\Sigma_1$ of all segments with beginning in $\mathcal{C}(X)$ are compact subsets of $\{2^X\}^E$, where $E$ is the unit interval. Now $\Sigma$ is an equicontinuous collection of mappings, for, for any segment $A_t$, we have $|\mu(A_{t'})-\mu(A_{t''})| = |t'-t''|(\mu(A_1)-\mu(A_0)) \leqq |t'-t''|$. Hence by 1.5, if $|t'-t''| < \eta(\epsilon)$ then $\rho^1(A_{t'}, A_{t''}) < \epsilon$. The relations 2.1, 2.2 clearly hold for any limit element and hence $\Sigma$ is compact. That $\Sigma_1$ is a closed subset of $\Sigma$ follows from the fact that for a convergent sequence of mappings, the limit of the beginning elements is the beginning element of the limit.

Let $A_t(x)$, for $x \in D$, be a continuous mapping of the set $D$ onto $\Sigma$ (or $\Sigma_1$). Now $A_t(x)$ is continuous simultaneously in $x$ and $t$, and since $A_1(x) = X$ for any $x \in D$, the mapping $f(x, y) = A_y(x)$ is a continuous mapping of the Cantor star onto $2^X$ (or $\mathcal{C}(X)$).

3. **Contractibility**(4). We now have the following lemma

3.1. LEMMA. *The following properties are equivalent*:

(a) *$\phi(X)$ is contractible in $2^X$.*

(b) *$2^X$ is contractible.*

(c) *$\mathcal{C}(X)$ is contractible (in itself).*

**Proof.** The proof is in three steps. First, (a) implies (b). If $\phi(X)$ is contractible in $2^X$ there exists a continuous mapping $F(a, t)$ of $X \times E$, where $E$ is the unit interval, into $2^X$, such that $F(a, 0) = \{a\}$, $F(a, 1) =$ a constant. Define for $A \in 2^X$, $\mathcal{J}(A, t) = \{F(a, t)\}$ for $a \in A$. Since $F(a, t)$ is a continuous mapping of $X \times E$ into $2^X$, $\mathcal{J}(A, t)$ maps continuously $2^X \times E$ into $2^{2^X}$. The deformation $\sigma(\mathcal{J}(A, t))$ is then continuous and contracts $2^X$ in itself.

Second, (b) implies (c). Suppose $2^X$ is contractible. There exists a mapping

(3) Actually, in order that a compact metric space $X$ be the continuous image of the Cantor star it is necessary and sufficient that there exist an equicontinuous family of mappings of $E$ into $X$ which includes a map of $E$ covering any pair of points. The proof of this proceeds exactly as that above.

(4) A space $X \subset Y$ is contractible in $Y$ if the identity transformation on $X$ is homotopic to a constant in $Y$.

$F(A, t)$ of $2^X \times E$ into $2^X$ such that $F(A, 0) = A$ and $F(A, 1) =$ a constant. Since $2^X$ is arcwise connected we can suppose $F(A, 1) = X$ for all $A \in 2^X$. Let $\mathcal{J}(A, t) = \{F(A, t')\}$ for $0 \leqq t' \leqq t$. Now $\mathcal{J}(A, t)$ is surely a continuous mapping of $2^X \times E$ into $2^{2^X}$. The deformation $G(A, t) = \sigma(\mathcal{J}(A, t))$ will then be continuous and will have the properties: $G(A, 0) = A$; $G(A, 1) = X$; if $0 \leqq t' < t'' \leqq 1$, then $G(A, t') \subset G(A, t'')$. Hence for $A$ fixed, $G(A, t')$, $0 \leqq t' \leqq t$ defines with proper parametrization, a segment from $G(A, 0)$ to $G(A, t)$. Hence by 2.6 if $A \in \mathcal{C}(X)$ then $G(A, t) \in \mathcal{C}(X)$ for every $t$, $0 \leqq t \leqq 1$. Hence $\mathcal{C}(X)$ is contractible.

Third, (c) implies (a). This is obvious.

**Remark.** It follows from the above arguments that if $2^X$ and $\mathcal{C}(X)$ are contractible then the deformation $G(A, t)$ can be chosen to satisfy

$$G(A + B, t) = G(A, t) + G(B, t).$$

If $0 \leqq t' < t'' \leqq 1$ then $G(A, t') \subset G(A, t'')$.

We shall consider spaces $X$ having the following property:

3.2. *For $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that if $a, b \in X$, $\rho(a, b) < \delta(\epsilon)$ and $a \in A \in \mathcal{C}(X)$, then there exists $B$, $b \in B \in \mathcal{C}(X)$ with $\rho^1(A, B) < \epsilon$.*

As a generalization of a theorem of Wojdyslawski (see [9]) we prove

3.3 THEOREM. *If $X$ has the property of 3.2 then $2^X$ and $\mathcal{C}(X)$ are contractible.*

**Proof.** In view of 3.1 it is sufficient to show that $\phi(X)$ is contractible in $2^X$. We define now a mapping of $X \times E$ into $2^{2^X}$ as follows: $\mathcal{J}(a, t) = \{A\}$ where $a \in A \in \mathcal{C}(X)$ and $\mu(A) = t$. Now for $G(a, t) = \sigma(\mathcal{J}(a, t))$ we have $G(a, 0) = \{a\}$ and $G(a, 1) = X$. Hence the proof reduces to showing the continuity of $\mathcal{J}(a, t)$.

First, for $a \in X$ we show uniform continuity in $t$. Suppose $0 \leqq t' \leqq t'' \leqq 1$. Then from 2.3 we see that for each $A_1 \in \mathcal{J}(a, t')$ there exists $A_2 \in \mathcal{J}(a, t'')$ such that $A_2 \supset A_1$, and similarly, given $A_2 \in \mathcal{J}(a, t'')$ we can find some $A_1 \in \mathcal{J}(a, t')$ with $A_2 \supset A_1$. Hence if $|t' - t''| < \eta(\epsilon)$ of 1.5 then every element of each of $\mathcal{J}(a, t')$ and $\mathcal{J}(a, t'')$ is within $\epsilon$ of some element of the other and $\rho^2(\mathcal{J}(a, t'), \mathcal{J}(a, t'')) < \epsilon$.

Finally, if $t$ is fixed $\mathcal{J}(a, t)$ is continuous in $a$. If $a$ and $b$ are near and $A \in \mathcal{J}(a, t)$ then by 3.2 we can choose $B$ near $A$, $b \in B \in \mathcal{C}(X)$. Now $\mu(B)$ is near $\mu(A)$. If $\mu(B) > \mu(A)$ we can choose $B_1$ on a segment from $\{b\}$ to $B$, $B_1$ near $A$ (see 1.5) with $\mu(B_1) = \mu(A)$. If $\mu(B) < \mu(A)$ we can choose $B_1$ on a segment from $B$ to $X$, with $\mu(B_1) = \mu(A)$. In either case we find $B_1$ near $A$, $B_1 \in \mathcal{J}(b, t)$, and continuity is demonstrated.

*Examples.* Let $X$ be the curve in the $xy$-plane defined by

$$\begin{aligned} y &= \sin \frac{1}{x}, && \text{for } 0 < x \leqq 1, \\ -1 \leqq y &\leqq 1, && \text{for } x = 0. \end{aligned}$$

It is easy to verify that condition 3.2 is satisfied for $X$ and hence $2^X$ and $\mathcal{C}(X)$ are contractible.

If we add to $X$ the interval

$$1 \leqq y \leqq \frac{3}{2}, \qquad \text{for} \quad x = 0,$$

then 3.2 is not satisfied for the curve $X_1$ so obtained. Nevertheless, since $X_1$ can be deformed into $X$, $2^{X_1}$ and $\mathcal{C}(X_1)$ are contractible. This shows that condition 3.2 is sufficient without being necessary.

If we now add to $X_1$ the points

$$y = \frac{1}{2} + \sin\frac{1}{x}, \qquad \text{for} \quad -1 \leqq x \leqq 0,$$

we obtain a curve $X_2$ for which $2^{X_2}$ and $\mathcal{C}(X_2)$ fail to be contractible. If $\mathcal{C}(X_2)$ were contractible we could suppose the deformation $F(A, t)$ satisfied the condition: If $0 \leqq t' < t'' \leqq 1$ then $F(A, t') \subset F(A, t'')$. If $a \in X_2$ and $a$ has a positive $x$-coordinate, there will exist $t_0$ such that $F(a, t_0) \subset X$ and $F(a, t_0)$ contains the interval $-1 \leqq y \leqq 1$ for $x = 0$. If $b \in X_2$ has a negative $x$-coordinate, every continuum containing $b$ is at least one-half unit from $F(A, t_0)$. But $a$ and $b$ can be chosen arbitrarily close, and we have a contradiction.

3.4. THEOREM. *The space $2^X$ is acyclic in all dimensions.*

**Proof.** Suppose $Z$ is a $\delta$-cycle in $2^X$, that is, an abstract cycle with vertices in $2^X$, with the diameter of every simplex less than or equal to $\delta$. For $A \in 2^X$ let $F(A)$ be the set of points in $X$ each of which is at most $\delta$ distance from some point of $A$. Now if $\rho^1(A, B) \leqq \delta$, then $\rho^1(F(A), F(B)) \leqq \delta$, for every point of $F(A)$ is at most $\delta$ distance from some point of $A$, and this point belongs to $F(B)$. Hence if we map each vertex $A_i$ of $Z$ into $F(A_i)$ we obtain a $\delta$-cycle $Z_1$ with each vertex at most $\delta$ from the corresponding vertex of $Z$. But from the definition of $F$ it follows that there is an integer $n$ such that the $n$th iteration of $F$ carries every $A \in 2^X$ into $X$. Hence $Z$ is $3\delta$ homologous to a cycle on $X \in 2^X$. The theorem follows.

**Remark.** In case $X$ satisfies the condition 3.2 then the preceding theorem as well as a similar theorem for $\mathcal{C}(X)$ is an obvious consequence of 3.3.

*Problem.* Is $\mathcal{C}(X)$ always acyclic in all dimensions?

4. **Local connectedness and retraction properties.** Before proceeding we note two lemmas:

4.1. LEMMA. *If $X$ is Peanian then $2^X$ and $\mathcal{C}(X)$ are contractible.*

**Proof.** Any Peano continuum surely has the property of 3.2.

4.2. LEMMA. *If $\mathcal{A}$ is a Peanian subset of $2^X$ (or $\mathcal{C}(X)$) then $A$ is contractible over a subset $\mathcal{B}$ of $2^X$ (or $\mathcal{C}(X)$) such that diameter $\mathcal{A}$ = diameter $\mathcal{B}$.*

**Proof.** If $\mathcal{J}(A, t)$ is a function contracting $\phi(\mathcal{A})$ in $\mathcal{C}(\mathcal{A})$, ($\mathcal{C}(\mathcal{A})$ is a subset of $\mathcal{C}(2^X)$), then $\sigma(\mathcal{J}(A, t))$ contracts $\mathcal{A}$ in $\sigma(\mathcal{C}(\mathcal{A}))$. Further, $\mathcal{C}(\mathcal{A})$ has the same diameter as $\mathcal{A}$, and $\sigma$ is a contraction.

The following local connectivity property implies local $p$-connectedness in the sense of Lefschetz(8) for $p>0$.

4.3. THEOREM. *Let $K$ be a finite complex, $K_1$ a subcomplex including all the 1-dimensional simplices of $K$, and $f(K_1)\subset 2^X$ (or $\mathcal{C}(X)$) a continuous mapping such that the partial image of any simplex of $K$ is of diameter less than $\epsilon$. Then $f$ may be extended to a mapping of all of $K$ into $2^X$ (or $\mathcal{C}(X)$) so that the diameter of the image of any simplex is less than $\epsilon$.*

**Proof.** First, let $f(S^n)\subset 2^X$ (or $\mathcal{C}(X)$), $n\geqq 1$, be a map of the surface of an $(n+1)$-cell $E^{n+1}$. Then, by 4.2, $f$ may be extended to a map of all of $E^{n+1}$ into $2^X$ (or $\mathcal{C}(X)$), for the image of $S^n$ is a Peano continuum. Now let $f(K_1)$ be the mapping given in the lemma. Then $\phi f$ is a map of $K_1$ into $\phi(2^X)\subset\mathcal{C}(2^X)$. We can now extend $f$ to all of each 2-simplex $x^2$ of $K$ so that $x^2$ maps into $\mathcal{C}(f(x^2\cdot K_1))$. Repeating this process, one dimension at a time, we arrive at a mapping $\bar{f}$ of all of $K$, identical with $\phi f$ on $K_1$, and such that the image of any simplex $x^n$ is contained in $\mathcal{C}(f(x^n\cdot K_1))$. Hence the diameter of $\bar{f}(x^n)$ equals the diameter of $f(x^n\cdot K_1)$. Since $\sigma$ is a contraction, we see that $\sigma\bar{f}$ is the required extension of $f$.

We now reprove a theorem of Wojdyslawski (see [10]; also [7] and [8]).

4.4. THEOREM (Wojdyslawski). *The following statements are equivalent:*

(a) *$X$ is Peanian.*
(b) *$2^X$ is Peanian.*
(b′) *$\mathcal{C}(X)$ is Peanian.*
(c) *$2^X$ is an absolute retract.*
(c′) *$\mathcal{C}(X)$ is an absolute retract*(9).

**Proof.** The proof is contained in the following three assertions:

First, (a) implies (b) and (b′). Suppose that any two points of $X$ less than $\nu(\epsilon)$ apart can be joined by a continuum of diameter less than $\epsilon$. Then if $A$, $B\in 2^X$, $\rho^1(A, B)<\nu(\epsilon)$, the set $C$ consisting of all points which can be joined to $A$ by continua of diameter at most $\epsilon$ has the properties: $\rho^1(A, C)\leqq\epsilon$, $\rho^1(B, C)\leqq 2\epsilon$, $A+B\subset C$ and every component of $C$ intersects both $A$ and $B$. Hence by 2.3 there exist segments $A_t$ and $B_t$ from $A$ to $C$ and $B$ to $C$, respectively. The continuum $\mathcal{A}=\{A_t\}+\{B_t\}$, $0\leqq t\leqq 1$, is of diameter less than or equal

(8) See S. Lefschetz, *Topology*, American Mathematical Society Colloquium Publications, vol. 12, 1930, p. 91.

(9) A space $X\subset Y$ is a *retract* of $Y$ if there exists a continuous transformation $f(Y)=X$ where $f$ is the identity on $X$. The metric separable space $X$ is an *absolute retract* if it is a retract of every metric space in which it can be imbedded. See K. Borsuk, *Sur les rétractes*, Fundamenta Mathematicae, vol. 17 (1931), pp. 152–170.

to $3\epsilon$, and $\mathcal{A}\subset\mathcal{C}(X)$ if $A$ and $B$ belong to $\mathcal{C}(X)$. Hence $2^X$ and $\mathcal{C}(X)$ are Peanian.

Second, (a) implies (c) and (c'). Combining the result of the previous paragraph with that of 4.3 we have: If $K$ is a finite complex, $K_0$ a subcomplex including all of the vertices of $K$, and if $f(K_0)\subset 2^X$ (or $\mathcal{C}(X)$) is a mapping such that the partial image under $f$ of any simplex of $K$ is of diameter less than $\nu(\epsilon/6)$, then $f$ can be extended to a mapping of all of $K$ into $2^X$ (or $\mathcal{C}(X)$) such that the image of any simplex of $K$ is of diameter at most $\epsilon$. This result, by a characterization of Lefschetz(7), implies that $2^X$ and $\mathcal{C}(X)$ are absolute retracts.

Third, either one of (b) or (b') implies (a). If $a, b\in X$, and $\phi(a)$ and $\phi(b)$ can be joined in $2^X$ by a continuum $\mathcal{A}$ of diameter $d$, then by 1.2 $\sigma(\mathcal{A})$ is a continuum in $X$ about $a+b$ of diameter at most $d$.

4.5. THEOREM. *Let $Y$ be a compact, locally connected subset of a metric space $Z$, and let $f(Y)$ be a continuous mapping of $Y$ into $2^X$ (or $\mathcal{C}(X)$). Then $f$ can be extended to a continuous mapping of all $Z$ into $2^X$ (or $\mathcal{C}(X)$).*

**Proof.** The set $f(Y)$ is locally connected, and since each hyperspace is arcwise connected, we can find a Peano continuum $\mathcal{A}$, $f(Y)\subset\mathcal{A}$, $\mathcal{A}$ in $2^X$ or $\mathcal{C}(X)$, respectively. Since $\mathcal{C}(\mathcal{A})$ is an absolute retract we can extend(8) the transformation $\phi f$ of $Y$ to a mapping $\bar{f}$ of $Z$ into $\mathcal{C}(\mathcal{A})$. The mapping $\sigma\bar{f}$ is then the required extension of $f$.

**Remark.** Consider any closed subset $\mathcal{A}$ of $2^X$ having the property: If $A\in\mathcal{A}$ and if $B\supset A$ and every component of $B$ intersects $A$ then $B\in\mathcal{A}$. All the results of §§2, 3, 4 for $2^X$ (except 3.4) can be shown by precisely the same reasoning to hold for such a set $\mathcal{A}$. In particular the space $\mathcal{C}_n(X)$ consisting of all closed subsets of $X$ having at most $n$ components, and the space $\mathcal{C}^d(X)$ consisting of all closed sets of diameter greater than or equal to $d$ have these stated properties of $2^X$.

5. **Dimension of hyperspaces.** Further topological properties are now obtained.

5.1. *The space $2^X$ always contains the homeomorph of the "fundamental cube."*

**Proof.** Choose $A_i\in\mathcal{C}(X)$, a sequence of nondegenerate disjoint continua tending to a point $a\notin A_i$ for any $i$. Now each $2^{A_i}$ contains a nondegenerate arc $B_i$ and $2^X$ contains topologically the infinite cartesian product $B_1\times B_2\times\cdots$. The theorem follows.

If $X$ is Peanian and $A\in\mathcal{C}(X)$ then the *order of $A$ in $X$* is the smallest integer $n$ such that there exists within any $V_\epsilon(A)$ a neighborhood of $A$ with

(7) Annals of Mathematics, (2), vol. 35 (1934), pp. 118–129.
(8) This is a property of absolute retracts. See Footnote 6.

boundary consisting of at most $n$ points. If no such integer exists then $A$ is said to be of *non-finite* order.

5.2. LEMMA. *If $X$ is Peanian the order of $A$ is finite for every $A \in \mathcal{C}(X)$ if and only if $X$ is a graph.*

**Proof.** We need only show that if $X$ is not a graph, $X$ contains a continuum of non-finite order. If $X$ contains no point constituting a continuum of non-finite order, $X$ must contain an infinite sequence $a_i$ of ramification points, and we can suppose $a_i \to a$. If there exists an arc containing infinitely many of the $a_i$ this arc is of non-finite order. Otherwise, we can choose infinitely many arcs $a_{i_n}a$, forming a null sequence and with each $a_{i_n}$ contained in only one arc of the sequence. Then $\sum_n a_{i_n}a$ is a continuum of non-finite order.

If $A$ is a closed subset of $X$, $\mathcal{C}(X, A)$ is the subset of $\mathcal{C}(X)$ consisting of continua which contain $A$. If $A \in \mathcal{C}(X)$ then $A \in \mathcal{C}(X, A)$. Also $\mathcal{C}(X, 0) = C(X)$.

5.3. LEMMA. *If $X$ is Peanian, then for every $A \in \mathcal{C}(X)$ we have order $A \leqq \dim_A \mathcal{C}(X, A)$.*

**Proof.** If $A \in \mathcal{C}(X)$ is of order $n$, then, using the $n$-Bogensatz(9), we can choose arcs $B_1, \cdots, B_n$, $B_i \cdot A = a_i$ and $(B_i - a_i)$ a collection of disjoint sets. To each $(t_1, t_2, \cdots, t_n) \in B_1 \times B_2 \times \cdots \times B_n$ assign the continuum $A + \sum_1^n a_i t_i$. This correspondence is a homeomorphism and the theorem is proved.

5.4. THEOREM. *If $X$ is Peanian then* $\dim \mathcal{C}(X) < \infty$ *if and only if $X$ is a linear graph.*

**Proof.** If $\dim \mathcal{C}(X)$ is finite then 5.3 and 5.2 imply that $X$ is a linear graph. The other half of the theorem is contained in the following sharper statement.

5.5. THEOREM. *If $X$ is a connected linear graph then*

$$\begin{aligned}\dim \mathcal{C}(X) &= \max_{A \in \mathcal{C}(X)} (\text{order } A)\\ &= 2 + \sum (\text{order } a - 2),\end{aligned}$$

*the last summation being extended over all points $a \in X$ such that order $a \geqq 2$.*

**Proof.** Let $A_1, A_2, \cdots, A_m$ be the collection of connected sub-graphs of $X$. With each $A_i$ there is associated the collection $\mathcal{A}_i$ of continua in $X$ for which $A_i$ is the maximal sub-graph. Clearly, $\mathcal{C}(X)$ is the sum of the $\mathcal{A}_i$. If the order of $A_i$ is $n$, then there are, say, $m$ 1-cells containing a single 0-cell of $A_i$ and $k$ 1-cells containing 2 0-cells of $A_i$, where $m+2k=n$. By the argument used in 5.3, we see that $\mathcal{A}_i$ is homeomorphic with the $F_\sigma$-set in $n$-space given by the inequalities $0 \leqq x_i < 1$ for $i = 1, \cdots, n$, $x_{2j-1} + x_{2j} < 1$ for $j = 1, \cdots, k$. Since $\mathcal{A}_i$ is an $F_\sigma$(10),

(9) See "$n$-Bogensatz," K. Menger, *Kurventheorie*, p. 216.

(10) See "Summensatz," K. Menger, *Dimensiontheorie*, p. 92.

$$\dim \mathcal{C}(X) \leqq \max_i (\dim \mathcal{A}_i) = \max_i (\text{order } A_i) \leqq \max_{A \in \mathcal{C}(X)} (\text{order } A).$$

The other necessary inequality is contained in 5.3.

The equality $\max_{A\in\mathcal{C}(X)}$ (order $A$) $=2+\sum$(order $a-2$) can be obtained by a simple induction argument.

**Remark.** If $X$ is a linear graph $\mathcal{C}(X)$ is actually a polyhedron. We have also the property: If $X$ is Peanian and $\mathcal{C}(X)$ has finite dimension at every one of its points then $\mathcal{C}(X)$ must have finite dimension.

**6. Local separating points.** In this section we prove a theorem of G. T. Whyburn.

6.1. THEOREM. *If $X$ is Peanian, $A$ any closed subset of $X$, $a\in X-A$, then $a$ is a local separating point of $X$ if and only if $\mathcal{C}(X, A+a)$ contains interior points relative to $\mathcal{C}(X, A)$.*

**Proof.** First, let $a$ be a nonlocal separating point, $a\in X-A$. For $B\in\mathcal{C}(X, A+a)$ and $\epsilon>0$ choose a connected neighborhood $U$ of $a$ of diameter less than $\epsilon$ so that $\overline{U}\cdot A=0$. Choose a neighborhood $V$ of $a$, $\overline{V}\subset U$, such that $U-V$ is connected. Then $(B+\overline{U-V})\in\mathcal{C}(X, A)-\mathcal{C}(X, A+a)$ and is at most $\epsilon$ distance from $B$. Hence $\mathcal{C}(X, A)-\mathcal{C}(X, A+a)$ is dense in $\mathcal{C}(X, A)$.

Second, let $a$ be a local separating point of $X$ and $U$ a connected neighborhood of $a$ such that $U-a=U_1+U_2$, $\overline{U}_1\cdot\overline{U}_2=a$. Let $V$ be a connected neighborhood of $a$ with $\overline{V}\subset U$. Choose a continuum $B\supset A+\overline{V}$ and intersecting the boundary of only one of $U_1$ and $U_2$ in points other than $a$. Any continuum sufficiently near $B$ intersects both $V\cdot U_1$ and $V\cdot U_2$ and fails to intersect the boundary of one of $U_1$ and $U_2$ in a point different from $a$. Hence $a$ is a point of this continuum and $B$ is interior to $\mathcal{C}(X, A+a)$ relative to $\mathcal{C}(X, A)$.

**Remark.** If $X$ is non-Peanian and $a$ is a local separating point then $\mathcal{C}(X, a)$ contains interior points relative to $\mathcal{C}(X)$. The converse is not necessarily true, however.

If $A$ is the null set we have this corollary.

6.2. COROLLARY. *If $X$ is Peanian, $a\in X$ then $a$ is a local separating point if and only if $\mathcal{C}(X, a)$ contains interior points relative to $\mathcal{C}(X)$.*

6.3. THEOREM (G. T. Whyburn[11]). *If $X$ is Peanian and $a_i\in X$ is a sequence of nonlocal separating points, then $X^*=X-\sum a_i$ is connected and locally connected.*

*In fact, if $b_1$, $b_2\in X^*$, and $b_1$ and $b_2$ can be joined in $X$ by a continuum of diameter less than $\epsilon$ then the same holds in $X^*$.*

**Proof.** The set $\prod_1^\infty(\mathcal{C}(X, b_1+b_2)-\mathcal{C}(X, b_1+b_2+a_n))$ is by the theorem of

(11) *Semi-closed sets and collections*, Duke Mathematical Journal, vol. 2 (1936), pp. 684–690. The above theorem is contained in Theorem 3.2 of the paper cited. I owe this proof to S. Eilenberg.

Baire, dense in $\mathcal{C}(X, b_1+b_2)$, since by 6.1 each set in the product is dense and open in $\mathcal{C}(X, b_1+b_2)$. Hence any continuum about $b_1+b_2$ is the limit of continua about $b_1+b_2$ in $X^*$. The theorem follows.

7. **Continuous transformations.** Here we show that for a continuous transformation $f(X)=Y$ we may under certain conditions find $X_0\subset X$, with $X_0$ closed and of dimension 0, such that $f(X_0)=Y$.

7.1. LEMMA. *If $f(E^2)=E^1$ is a continuous mapping of the unit square onto the unit interval, then there exist two disjoint arcs ab and cd in $E^2$, each containing at most one boundary point of $E^2$, such that $f(ab+cd)=E^1$.*

**Proof.** The interior of $E^2$ maps into a connected set which is dense in $E^1$. Choose $a\in f^{-1}(0)$, $b\in f^{-1}(2/3)$, $c\in f^{-1}(1/3)$, $d\in f^{-1}(1)$ so that $b$ and $c$ do not belong to the boundary $E^2$. Choose $ab$ and $cd$ disjoint arcs in $E^2$ having at most $a$ and $d$ in common with the boundary of $E^2$. Then $f(ab)\supset(0, 2/3)$ and $f(cd)\supset(1/3, 1)$.

7.2. THEOREM[12]. *If $f(E^2)=E^1$ is a continuous mapping of the unit square onto the unit interval then there exists a closed totally disconnected subset $Z$ of $E^2$ such that $f(Z)=E^1$.*

**Proof.** Let $\mathcal{A}$ be the subset of $2^{E^2}$ consisting of all subsets of $E^2$ which map onto $E^1$ under $f$. Let $\mathcal{A}_\epsilon$ be the subset of $\mathcal{A}$ consisting of sets having only components of diameter less than $\epsilon$. Clearly $\mathcal{A}_\epsilon$ is open in $\mathcal{A}$, and we shall show $\mathcal{A}_\epsilon$ is dense in $\mathcal{A}$. Since a residual set in a complete space is nonvacuous, it will be true that $\prod\mathcal{A}_{1/n}\neq 0$, and any $A\in\prod\mathcal{A}_{1/n}$ will be a totally disconnected closed set mapping on $E^1$.

Suppose $A\in\mathcal{A}$ and $\epsilon>0$ are given. We shall find $B\in\mathcal{A}_\epsilon$, $\rho^1(A, B)<\epsilon$. Choose a subdivision of $E^2$ into closed squares $S_1, S_2, \cdots, S_q$, each of diameter less than $\epsilon/4$. For each $S_p$ which intersects $A$ choose arcs $a_pb_p$ and $c_pd_p$ by 7.1, each mapping onto $f(S_p)$, and let $B$ be the sum of the arcs so chosen. Since dia $S_p<\epsilon/4$, $B$ has only components of diameter less than $\epsilon$. Since $B$ intersects those and only those squares $S_p$ which are cut by $A$, $\rho^1(A, B)<\epsilon$ and $f(B)\supset E^1$. Hence $B\in\mathcal{A}_\epsilon$ and the proof is complete.

We now obtain a similar theorem with more general space and more special type of transformation. First, consider a transformation $f(X)=Y$ where

7.3. (a) *$X$ is compact and metric and* dim $Y<\infty$.

(b) *$f$ is monotone and interior*[13].

(c) dia $f^{-1}(y)>0$ *for all* $y\in Y$.

---

(12) I owe this theorem to S. Eilenberg and L. Zippin.

(13) A transformation is *monotone* if the inverse of every point in the image space is connected. See R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, 1932, chap. 5. The term "monotone" is due to C. B. Morrey, American Journal of Mathematics, vol. 57 (1935), pp. 17–50. A transformation is interior if open sets map into open sets. For references see G. T. Whyburn, Duke Mathematical Journal, vol. 3 (1937), pp. 370–381.

7.4. Lemma. *Under the hypothesis of* 7.3 *for any* $A\in 2^X$ *where* $f(A)=Y$ *and for any* $\epsilon>0$ *there exists* $B\in 2^X$ *such that*

(a) $\rho^1(A,B)<\epsilon$.

(b) $f(B)=Y$.

(c) *Every component of* $B$ *is of diameter less than* $\epsilon$.

**Proof.** It is sufficient to find $B\subset V_\epsilon(A)$ and satisfying (b) and (c) since by adding a finite number of points to such a $B$ we may obtain a set within $\epsilon$ of $A$. Let $V_0=V_{\epsilon/2}(A)$. We shall need these three lemmas:

7.5. Lemma. *There exists* $\tau(\epsilon)>0$ *such that* $f(V_\epsilon(x))\supset V_{\tau(\epsilon)}(f(x))$ *for every* $x\in X$.

7.6. Lemma. *There exists* $d>0$ *such that for any* $y\in Y$ *there is a component* $A_y$ *of* $V_0\cdot f^{-1}(y)$ *such that* dia $A_y\geqq d$.

7.7. Lemma. *There exists an integer* $N$ *such that* $Y$ *allows an arbitrarily fine covering by open sets,* $W_1,\cdots,W_m$ *such that at most* $N$ *of the sets* $\overline{W}_i$ *intersect any given* $\overline{W}_r$.

The first of these is a simple consequence of interiority, the second follows since dia $f^{-1}(y)>0$ for all $y\in Y$, and the third is true since $Y$ can be imbedded in a finite-dimensional euclidean space.

Let $s=\min\,[\epsilon/3,\ d/8N]$ and construct a covering of $Y$ of the type 7.7 with dia $W_r<\tau(s)$ for $r=1,\cdots,m$. Let $U_i=f^{-1}(W_i)$. Choose $a_1\in U_1\cdot V_0$ and let $A_1=\overline{U_1\cdot V_s(a_1)}$. Choose successively then $a_r\in U_r\cdot V_0$ and $A_r=\overline{U_r\cdot V_s(a_r)}$ so that $A_r\cdot A_i=0$ for $i<r$. That this is always possible is shown as follows: Choose $y\in W_r$ and $A_y$ of 7.6. At most $N$ of the sets $A_1\cdots,A_{r-1}$ intersect $\overline{U}_r$, and each $A_i$ is of dia less than $2s$. If $\sum_1^{r-1}V_s(A_i\cdot\overline{U}_r)$ intersected $V_s(a)$ for every $a\in A_y$ then $\sum_1^{r-1}V_{2s}(V_s(A_i\cdot\overline{U}_r))\supset A_y$ and dia $A_y\leqq N\cdot 8s<d$ which is impossible. Hence it is possible to choose $A_1,\cdots,A_m$ as prescribed. Finally, $f(A_r)\supset W_r$ and $\sum A_r\subset V_\epsilon(A)$. Let $B=\sum A_r$ and the result follows.

7.8. Theorem. *Let* $f(X)=Y$ *be a monotone interior transformation of a compact metric space* $X$ *into a set* $Y$ *of finite dimension. Then there exists a closed totally disconnected subset* $X_0$ *of* $X$ *mapping onto* $Y$ *if and only if the set of points on which* $f$ *is* 1-1 *is a totally disconnected subset of* $Y$.

**Proof.** First, suppose $f^{-1}(y)$ contains more than a single point for every $y\in Y$. If $\mathcal{A}\subset 2^X$ is the set of all sets mapping onto $Y$ under $f$, then by 7.4 the subset $\mathcal{A}_{1/n}$ of sets with components of diameter less than $1/n$ is dense in $\mathcal{A}$. Any set belonging to the residual set $\prod\mathcal{A}_{1/n}$ then satisfies the theorem.

Second, suppose $f^{-1}(y)$ consists of a single point for all $y\in B$, $B$ a totally disconnected set. Since $f$ is interior, $B$ is closed. Let $V_n=V_{1/n}(B)$ and using the result of the previous paragraph choose $A_n$, closed, totally disconnected and mapping on $\overline{V}_n-V_{n+1}$. Then $X_0=\sum A_n+f^{-1}(B)$ is easily seen to be totally disconnected and maps onto $Y$.

Finally, if $f$ is 1-1 on a continuum, it is clearly impossible to find $X_0$ satisfying the theorem.

8. **Knaster continua.** A compact metric continuum is *indecomposable* if it cannot be written as the sum of two proper subcontinua.

8.1. LEMMA. *If $X$ is indecomposable and $\mathcal{A}_{AB}$ is an arc in $\mathcal{C}(X)$ with $\sigma(\mathcal{A}_{AB})=X$ then $X\in\mathcal{A}_{AB}$.*

**Proof.** Let $C$ be the first element in order from $A$ to $B$ such that $\sigma(\mathcal{A}_{AC})=X$. For each $C_1$ preceding $C$ the continuum $\sigma(\mathcal{A}_{AC_1})$ is contained in the composant about $A$ of $X$, and hence $\sigma(\mathcal{A}_{C_1C})$ contains points both in this composant and in its complement. Thus $\sigma(\mathcal{A}_{C_1C})=X$ for all $C_1$ preceding $C$, and therefore $C=X$.

8.2. THEOREM. *In order that X be indecomposable it is necessary and sufficient that $\mathcal{C}(X)-X$ fail to be arcwise connected.*

**Proof.** If $X$ is indecomposable then for any arc $\mathcal{A}_{AB}$ where $A$ and $B$ lie in different composants of $X$ we have $\sigma(\mathcal{A}_{AB})=X$ and hence $X\in\mathcal{A}_{AB}$. Thus $\mathcal{C}(X)-X$ is not arcwise connected.

If $X$ is not indecomposable write $X=A_1+A_2$, $A_i\in\mathcal{C}(X)$, $A_i\neq X$ for $i=1, 2$. If $B\in\mathcal{C}(X)$, $B\neq X$, and $a\in B\cdot A_1\cdot A_2$ then there exists a segment joining $\{a\}$ to $B$, and also segments joining $\{a\}$ to both $A_1$ and $A_2$. If $B\subset A_1$ there is a segment from $B$ to $A_1$. In any event $B$ can be joined by an arc to both of $A_1$ and $A_2$ in $\mathcal{C}(X)-X$ and the theorem is proved.

A compact metric continuum is a *Knaster continuum*(14) if every subcontinuum is indecomposable. If $X$ is a Knaster continuum and if $A$, $B\in\mathcal{C}(X)$, then either $AB=0$, $A\supset B$ or $B\supset A$. Hence:

8.3. LEMMA. *If $X$ is a Knaster continuum, $A$, $B\in\mathcal{C}(X)$, $AB\neq 0$ and $\mu(A)=\mu(B)$ then $A=B$.*

8.4. THEOREM. *The continuum $X$ is a Knaster continuum if and only if $\mathcal{C}(X)$ contains a unique arc between every pair of its elements.*

**Proof.** If $\mathcal{C}(X)$ contains a unique arc between every pair of elements then for any $A\in\mathcal{C}(X)$, $\mathcal{C}(A)-A$ must fail to be arcwise connected and hence by 8.2 indecomposable. Therefore $X$ is a Knaster continuum.

Suppose $X$ is a Knaster continuum and $\mathcal{A}_{AB}$ an arc in $\mathcal{C}(X)$. Since $\sigma(\mathcal{A}_{AB})$ is indecomposable by 8.1 we have $\sigma(\mathcal{A}_{AB})\in\mathcal{A}_{AB}$. Hence the function $\mu$ assumes a unique maximum on any simple arc, and if $C=\sigma(\mathcal{A}_{AB})$ then $\mu$ must be strictly monotone on each of $\mathcal{A}_{AC}$ and $\mathcal{A}_{CB}$. For $C_1\in\mathcal{A}_{AC}$ we then have $C_1=\sigma(\mathcal{A}_{AC_1})$. It follows that $\mathcal{A}_{AC}$ and $\mathcal{A}_{CB}$ are, with proper parametrization,

(14) The only known example of a continuum of this type was given by B. Knaster in his dissertation, *Un continu dont tout sous-continu est indécomposable*, Fundamenta Mathematicae vol. 3 (1922), pp. 247–286.

segments. From 8.3 we see that there exists a unique continuum containing $A$ at which $\mu$ assumes any specified value. Hence the arc $\mathcal{A}_{AB}$ is unique.

8.5. THEOREM. *If $X$ is a Knaster continuum, for every $\epsilon>0$ there exists a monotone interior transformation $f(X)=Y$ such that $0<\text{dia } f^{-1}(y)<\epsilon$ for all $y\in Y$.*

**Proof.** Choose $d>0$ such that if $\mu(A)=d$ then dia $A<\epsilon$. For each $a\in X$ there is, by 8.4, a unique $A(a)\in\mathcal{C}(X)$ such that $a\in A(a)$ and $\mu(A(a))=d$. If $A(a)\cdot A(b)\neq 0$ then by 8.3 $A(a)=A(b)$. If lim $a_i=a$, then since $\mu$ is continuous and $A(a)$ single-valued lim $A(a_i)=A(a)$. The map $A(a)$ is then a continuous monotone interior transformation of $X$ into $\mathcal{C}(X)$ and satisfies the conditions of the theorem.

8.6. THEOREM. *If $X$ is a Knaster continuum and if there exists, for every $\epsilon>0$, a monotone interior transformation $f(X)=Y$ such that:*

(a) $0<\text{dia } f^{-1}(y)<\epsilon$ *for all* $y\in Y$;

(b) dim $Y<\infty$,

*then* dim $X=1$.

**Proof.** Under the hypotheses of the theorem we shall exhibit an $\epsilon$-covering of order 2 of $X$ by closed sets. Choose $X_0\subset X$, by 7.8, closed, totally disconnected, with $f(X_0)=Y$. Let $U$ be an open set about $X_0$ so that the diameter of any component of $\overline{U}$ is less than $\epsilon$. Every component of $X-U$ is of diameter less than $\epsilon$, for if $A\subset X-U$, dia $A\geqq\epsilon$ then for $a\in A$, $f^{-1}(f(a))\subset A$. But this contradicts the fact that $f(X_0)=Y$. Write each of $\overline{U}$ and $X-U$ as the sum of a finite number of closed disjoint sets of diameter less than $\epsilon$. The resulting covering of $X$ is surely of order 2.

From 8.4 and 8.6 and the fact that the monotone image of a Knaster continuum is also a Knaster continuum we have:

8.7. THEOREM. *If $X$ is a Knaster continuum of dimension greater than* 1 *then:*

(a) *for every $\epsilon>0$ there is a monotone interior transformation $f(X)=Y$ $0<\text{dia } f^{-1}(y)<\epsilon$ for all $y\in Y$, with* dim $Y=\infty$;

(b) *there exists an $\epsilon>0$ such that for any monotone interior $f(X)=Y$, with $0<\text{dia } f^{-1}(y)<\epsilon$ for all $y\in Y$, it is true that* dim $Y=\infty$;

(c) *there exist Knaster continua of infinite dimension.*

**Remark.** Theorem 8.6 could be demonstrated without the restriction (b) on dimension if instead of 7.8 we had at at our disposal the theorem: *If $f(X)=Y$ is monotone interior then there exists $X_0$, closed in $X$, with $f(X_0)=Y$, such that $X_0\cdot f^{-1}(y)$ is totally disconnected for all $y\in Y$; that is, such that $f$ is light on $X_0$.* This statement is much weaker, except for restriction on dim $Y$, than 7.8, and its truth would imply that every Knaster continuum is of dimension one.

## Bibliography

1. K. Borsuk and S. Mazurkiewicz, *Sur l'hyperespace d'un continu*, Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, vol. 24 (1931), pp. 149–152.

2. K. Kuratowski, *Topologie*, p. 92.

3. S. Mazurkiewicz, *Sur les continus absolument indécomposables*, Fundamenta Mathematicae, vol. 16 (1930), pp. 151–159.

4. ———, *Sur l'hyperespace d'un continu*, Fundamenta Mathematicae, vol. 18 (1932), pp. 171–177.

5. ———, *Sur le type C de l'hyperespace d'un continu*, Fundamenta Mathematicae, vol. 20 (1933), pp. 52–53.

6. ———, *Ein Zerlegungsatz*, Fundamenta Mathematicae, vol. 23 (1934), pp. 11–14.

7. L. Vietoris, *Kontinua zweiter Ordnung*, Monatshefte für Mathematik und Physik, vol. 33 (1923), pp. 49–62.

8. T. Wazewski, *Sur un continu singulier*, Fundamenta Mathematicae, vol. 4 (1923), pp. 214–235.

9. M. Wojdyslawski, *Sur la contractilité des hyperespaces de continus localement connexes*, Fundamenta Mathematicae, vol. 30 (1938), pp. 247–252.

10. ———, *Rétractes absolus et hyperespaces des continus*, Fundamenta Mathematicae, vol. 32 (1939), pp. 184–192.

University of Virginia,
Charlottesville, Va.
University of Notre Dame,
Notre Dame, Ind.

# TOPICS IN THE THEORY OF MARKOFF CHAINS

BY

J. L. DOOB

Let $P(t)$: $(p_{ij}(t))$ be a matrix (finite- or infinite-dimensional), depending on $t>0$, whose elements satisfy the following conditions

$$(0.1) \qquad p_{ij}(t) \geqq 0, \quad \sum_j p_{ij}(t) = 1, \quad P(s)P(t) = P(t)P(s) = P(s+t).$$

Then $p_{ij}(t)$ can be considered a transition probability of a Markoff chain: A system is supposed which can assume various numbered states, and $p_{ij}(t)$ is the probability that the system is in the $j$th state at the end of a time interval of length $t$, if it was in the $i$th state at the beginning of the interval. The present paper will be divided into two parts. In the first, the regularity properties of $P(t)$, and its asymptotic properties as $t\to 0$, $t\to\infty$ are studied. These problems have been solved in the finite-dimensional case by Doeblin(1). In the infinite-dimensional case new situations can arise, and the results are somewhat different. The method of approach is new, depending on two theorems (Theorems 2 and 3) concerning matrices whose elements are non-negative, and which have row sums less than or equal to 1. The method of approach can also be applied to the study of the asymptotic properties of the powers of a matrix of non-negative elements, with row sums 1. In the second part of the paper, the actual transitions connected with Markoff chains are investigated: That is, the properties of the function $\xi(t)$, the number of the state which the given system assumes at time $t$, are investigated. The continuity properties of $\xi(t)$ are analyzed, and related to the regularity properties of the $p_{ij}(t)$.

LEMMA 1. *Suppose that the function $f(t)$, defined for all $t>0$, satisfies the functional equation*

$$(1.1) \qquad f(s+t) = \sum_n g_n(s)h_n(t) \qquad (s, t > 0),$$

*where $g_n(s)$, $h_n(s)$ are defined for $s>0$, where $h_n(s)$ is measurable, and where for each fixed $s$, if $0<a<b$, the series converges uniformly for $a\leqq t\leqq b$. Then $f(t)$ is continuous, for all $t>0$.*

---

Presented to the Society, December 31, 1941; received by the editors May 24, 1941.

(1) Bulletin des Sciences Mathématiques, (2), vol. 62 (1938), pp. 21–32, and vol. 63 (1939), pp. 35–37. In the following, these papers will be referred to as Doeblin (I). Fréchet has discussed the solutions of (0.1) in great detail, in the finite-dimensional case, with full references to earlier authors in his book *Traité du Calcul des Probabilités et de ses Applications*, vol. 1, Part 3, Book 2, *Méthode des fonctions arbitraires* · · · , Paris, 1938.

It will be sufficient to prove that if a value $t_0$ of $t$ is given, and if $\{\delta_j\}$ is a sequence of numbers approaching 0, then $f(t_0+\delta_{a_j})\to f(t_0)$, for some subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$. By a theorem of Auerbach[2], there is, corresponding to each $h_n(t)$, a subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$, such that

$$\lim_{j\to\infty} h_n(t+\delta_{a_j}) = h_n(t) \tag{1.2}$$

for almost all $t$ in the interval $0<t<t_0$. There is then, using the diagonal process, a subsequence $\{\delta_{a_j}\}$ of $\{\delta_j\}$ such that (1.2) is true for all $n$, $0<t<t_0$, except possibly on a $t$-set of measure 0. If $0<t<t_0$, and if $j$ is large,

$$f(t_0+\delta_{a_j}) = \sum_n g_n(t_0-t)h_n(t+\delta_{a_j}), \tag{1.3}$$

and if $t$ is not in the exceptional set, (1.3) implies, when $j\to\infty$,

$$f(t_0+\delta_{a_j}) \to \sum_n g_n(t_0-t)h_n(t) = f(t_0) \tag{1.4}$$

(because of the uniform convergence of the series in (1.3) with respect to $j$), as was to be proved.

THEOREM 1. *If the matrix function $P(t)$ satisfies* (0.1), *the measurability of the $p_{ij}(t)$ implies their continuity.*

This follows at once from Lemma 1. It has been shown by Doeblin (I) and it will be a corollary of results to be proved below, that the $p_{ij}(t)$ satisfying (0.1) are always continuous if the matrix $P(t)$ is finite-dimensional, even if measurability is not assumed. The following example shows that there are non-measurable solutions of (0.1).

In this example, the $p_{ij}(t)$ take on only the values 0, 1, and $P(t)$ is a permutation matrix. Hamel has shown that there is a function $f(x)$, defined for all real $x$, taking on only rational values, and satisfying the functional equation[3] $f(x+y)=f(x)+f(y)$. Let $\{r_n\}$ be an enumeration of all the rational numbers, and let $T_s$ be the transformation of these numbers taking $r_j$ into $r_j+f(s)$. The transformation can be represented by a matrix $P(s)$: $(p_{ij}(s))$, where $p_{ij}(s)=1$ if $T_s r_i=r_j$, and $p_{ij}(s)=0$ otherwise. Then evidently $P(s+t)=P(s)P(t)$, and (0.1) is satisfied. The functions $p_{ij}(t)$ are not measurable, since they obviously are not continuous.

The following theorem describes completely the solutions of (0.1) which are independent of $t$. It will be useful to weaken (0.1) slightly. The theorem is essentially known, at least in an indirect form[4].

---

(2) Fundamenta Mathematicae, vol. 11 (1928), pp. 196–197.

(3) Mathematische Annalen, vol. 60 (1905), pp. 459–462. To ensure that Hamel's $f(x)$ take on only rational values, we can set, using his notation, $f(a)=1$, $f(b)=\cdots=0$.

(4) Cf., for example, K. Yosida and S. Kakutani, Japanese Journal of Mathematics, vol. 16 (1939), pp. 47–55.

THEOREM 2. *Let $U$: $(u_{ij})$ be a matrix of elements satisfying the following conditions*:

$$u_{ij} \geqq 0, \qquad \sum_j u_{ij} \leqq 1, \qquad U^2 = U. \tag{2.1}$$

*Then the subscripts can be divided into mutually exclusive classes*([5]) $F, G_1, G_2, \cdots$ *such that*

(a) $u_{ij}=0$, *if* $j \in F$;

(b) *there are positive numbers* $u_j$, $j \notin F$, *such that*([6])

$$u_{ij} = \delta_{IJ} u_j, \qquad (\textit{if } i \in I, j \in J),$$

$$\sum_{j \in J} u_{ij} = \sum_{j \in J} u_j = 1 \qquad (i \in I = J);$$

(c) *there are non-negative numbers* $\{\rho_{iJ}\}$ *such that*

$$u_{ij} = \rho_{iJ} u_j \qquad (\textit{if } i \in F, j \in J).$$

*Conversely, if the* $u_{ij}$ *satisfy* (a), (b), (c), *then* (2.1) *is true.*

Suppose (2.1) is true. Define $F$ as the set of integers $j$ with $p_{ij}=0$ for all $i$. Then (a) is true. Unless $U$ is the null matrix, there will be subscripts not in $F$. The extreme members of the inequality

$$\sum_k u_{ik} = \sum_{j,k} u_{ij} u_{jk} = \sum_j u_{ij} \left( \sum_k u_{jk} \right) \leqq \sum_j u_{ij} \tag{2.2}$$

are equal; so if $\sum_k u_{jk} < 1$, $u_{ij} = 0 : j \in F$. Let $\xi_1, \xi_2, \cdots$ be any numbers satisfying the conditions:

$$\sum_j |\xi_j| < \infty, \qquad \sum_i \xi_i u_{ij} = \xi_j \qquad (\text{all } j). \tag{2.3}$$

Then $\xi_j = 0$ if $j \in F$. If $G$ is the set of integers $j$ for which $\xi_j > 0$,

$$\sum_{j \in G} \xi_j = \sum_{j \in G} \sum_i \xi_i u_{ij} \leqq \sum_{i,j \in G} \xi_i u_{ij} \leqq \sum_{i \in G} \xi_i, \tag{2.4}$$

and there is an impossible inequality unless $u_{ij}=0$ whenever $\xi_i<0$, $\xi_j>0$. Let $i, j$ be any two distinct integers not in $F$. Unless the $i$th and $j$th columns of $U$ are proportional (neglecting elements in the columns whose first subscripts are in $F$), there are integers $r, s$ $(\notin F)$, such that

---

([5]) The class $F$ may be absent, or the $G_\nu$ may be absent. The latter case will arise when and only when $U$ is the null matrix.

([6]) In the following, capital letters $I$, $J$, $K$ will be used to denote the $G_\nu$, and a subscript $i$ will always belong to the class $I$, and so on, unless the contrary is explicitly stated. The notation $\delta_{IJ}$ is the usual Kronecker $\delta$.

$$(2.5)\qquad \begin{vmatrix} u_{ri} & u_{rj} \\ u_{si} & u_{sj} \end{vmatrix} \neq 0.$$

Then $\lambda$, $\mu$ can be chosen so that $\lambda u_{ri}+\mu u_{si}<0$, $\lambda u_{rj}+\mu u_{sj}>0$. Since $\xi_k=\lambda u_{rk}+\mu u_{sk}$ provides a solution of (2.3) with $\xi_i<0$, $\xi_j>0$, it follows that $u_{ij}=0$ if (2.5) is true. The subscripts not in $F$ fall into classes, $G_1, G_2, \cdots$, putting subscripts in the same class if the corresponding partial columns are proportional; if $i, j$ ($\notin F$) are not in the same class, $u_{ij}=0$. The elements $u_{ij}$ with $i$ in a class $G_\nu$ determine a matrix of rank 1. The rows of this matrix are therefore proportional, in fact identical, since the row sums are 1. We can thus write $u_{ij}=\delta_{IJ}u_j$ ($i\in I, j\in J$), and

$$(2.6)\qquad \sum_{j\in J} u_j = \sum_i u_{j_0 i} = 1 \qquad (j_0 \in J).$$

If $i\in F, j\in J$,

$$(2.7)\qquad u_{ij} = \sum_k u_{ik}u_{kj} = \left(\sum_{k\in J} u_{ik}\right)u_j = \rho_{iJ}u$$

where $\rho_{iJ}$ is defined by the sum in the parentheses. If $j\notin F$, $u_j$ cannot vanish, since the elements of the $j$th column ($j\notin F$) cannot all vanish. We have now shown that (2.1) implies (a), (b), (c). Conversely, if (a), (b), (c) are true, (2.1) can be checked at once.

THEOREM 3. *Let $\mathfrak{M}$ be a set of matrices (finite- or infinite-dimensional) with non-negative elements, and row sums less than or equal to* 1. *Suppose that the matrices in $\mathfrak{M}$ form a group $\mathfrak{M}'$. There is then a $U\in\mathfrak{M}$ (the identity in $\mathfrak{M}'$) with $U^2=U$. If $U$ is the null matrix, $U$ is the only matrix in $\mathfrak{M}$, and $\mathfrak{M}'$ consists only of the identity. If $U$ is not the null matrix, we shall use the notation of Theorem* 2 *to describe its elements. The group $\mathfrak{M}'$ is always isomorphic to a permutation group acting on the $G_\nu$. If $(p_{ij})$ is a matrix of $\mathfrak{M}$, and if the corresponding permutation takes $I_1$ into $I_2$*[7], *then*

$$(3.1)\qquad \begin{aligned} p_{ij} &= \delta_{I_2J}u_j && (i \in I_1, j \in J),\\ p_{ri} &= \rho_{rI_1}u_i && (r \in F, i \in I_2),\\ p_{ij} &= 0 && (j \in F). \end{aligned}$$

Evidently if $U$ is the null matrix, it is the only matrix in $\mathfrak{M}$, and $\mathfrak{M}'$ consists only of the identity. We shall assume from now on that $U$ is not the null matrix, and use the notation of Theorem 2. Suppose that $P:(p_{ij})\in\mathfrak{M}$. Then since $P=PU=UP$,

[7] As usual, letters $I$, $J$ refer to the $G_\nu$.

$$p_{ik} = \sum_j p_{ij} u_{jk} = \sum_j u_{ij} p_{jk}. \tag{3.2}$$

If $k \in F$, (3.2) shows that $p_{ik} = 0$, the last equation of (3.1). If $i \in I$, $k \in K$, (3.2) becomes

$$p_{ik} = \left( \sum_{r \in K} p_{ir} \right) u_k, \tag{3.2'}$$

$$p_{ik} = \sum_{r \in K} u_r p_{rk}. \tag{3.2''}$$

According to (3.2′), $p_{ik}/u_k$ depends only on $i$, $K$, and according to (3.2″), $p_{ik}$ depends only on $I$, $k$. Then $p_{ik}/u_k$ depends only on $I$, $K$:

$$p_{ik} = \sigma_{IK} u_k \qquad (i \in I,\ k \in K). \tag{3.3}$$

There is a $P' : (p'_{ij})$ in $\mathfrak{M}$ which is the inverse of $P$ in $\mathfrak{M}'$. If we write $p'_{ik} = \sigma'_{IK} u_k$, for $i, j \notin F$, the equation $U = P'P$ implies

$$u_{ik} = \delta_{IK} u_k = \sum_J \sigma'_{IJ} \sigma_{JK} u_k. \tag{3.4}$$

The $\sigma_{IJ}$, $\sigma'_{IJ}$ are non-negative and

$$\sum_J \sigma'_{IJ} = \sum_k p'_{ik} \leqq 1. \tag{3.5}$$

If $I = K$ in (3.4), we obtain

$$1 = \sum_J \sigma'_{IJ} \sigma_{JI} \leqq \sum_J \sigma'_{IJ} \leqq 1. \tag{3.6}$$

There must be equality throughout in (3.6); therefore if $\sigma_{JI} < 1$, it follows that $\sigma'_{IJ} = 0$. The matrices $(\sigma_{IJ})$, $(\sigma'_{IJ})$ play symmetric roles; so if $\sigma'_{JI} < 1$, it follows that $\sigma_{IJ} = 0$. Then if $\sigma_{IJ} < 1$, $\sigma'_{JI} = 0 < 1$; so $\sigma_{IJ} = 0$. Each element in the matrix $(\sigma_{IJ})$ is either 1 or 0, and by (3.6) there is a 1 in each row of $(\sigma'_{IJ})$ and therefore in each row of $(\sigma_{IJ})$. If $\sigma_{I_1 I_2} = 1$, the matrix $(\sigma_{IJ})$ defines the permutation of the $G$, taking $I_1$ into $I_2$. The matrix $(\sigma'_{IJ})$ defines the inverse of this permutation. Equation (3.3) becomes the first equation of (3.1), equation (3.2) implies the second equation of (3.1), and the third equation of (3.1) has already been verified. The equations of (3.1) induce an isomorphism between the permutations defined by the $(\sigma_{IJ})$ permutation matrices and $\mathfrak{M}'$.

COROLLARY 1. *Suppose in Theorem 3 that $\mathfrak{M}$ contains its limit matrices*(8). *Then the corresponding permutation group on the $G_i$ has the property that each $G_j$*

(8) The matrices $\{M^{(n)} : (m_{ij}^{(n)})\}$ will be said to converge to $M : (m_{ij})$, $M^{(n)} \to M$, if $m_{ij}^{(n)} \to m_{ij}$ for all $i$, $j$. The limit matrices of $\mathfrak{M}$ are matrices which are limits of convergent sequences of matrices in $\mathfrak{M}$.

*can go only into a finite number of the* $G_\nu$. *If in addition it is supposed that corresponding to each* $A \in \mathfrak{M}$ *and positive integer n there is a* $B \in \mathfrak{M}$ *such that* $B^n = A$, *then* $\mathfrak{M}$ *consists of only a single matrix, of the type described in Theorem* 2.

Suppose that $\mathfrak{M}$ contains its limiting matrices, and that some $G_\nu$, say $G_\alpha$, goes into infinitely many $G_\nu$ under the permutations of the group. Then there is a limiting matrix $(p_{ij})$ of $\mathfrak{M}$ such that $p_{ij}=0$ if $i \in G_\alpha$. But a matrix with these rows of zeros cannot be in $\mathfrak{M}$, so $G_\alpha$ cannot have the supposed property. The first part of the corollary is thus proved. Now suppose both hypotheses of the corollary are satisfied. It will be sufficient to prove that the group of permutations on the $G_\nu$ is the identity. Let $G_\alpha$ be any $G_\nu$. We have already shown that $G_\alpha$ can go only into a finite number of $G_\nu$, say $G_{\alpha_1}, \cdots, G_{\alpha_j}$, under the permutations of the group. The permutations then permute $G_{\alpha_1}, \cdots, G_{\alpha_j}$ among themselves, and any element of the group of permutations on $G_{\alpha_1}, \cdots, G_{\alpha_j}$ has order a factor of $j!$. But any element in this group of permutations is by hypothesis the $j!$th power of some other element; it must therefore be the identity. Then $j=1$, and $G_\alpha$ is transformed into itself by every permutation of the group, as was to be proved.

Corollary 2. *Any matrix function* $P(t):(p_{ij}(t))$ *with measurable elements* $p_{ij}(t)$ *satisfying* (0.1) *for all t (including* 0 *and negative values) is independent of* $t: P(t) \equiv U$, *where U has the properties described in Theorem* 2.

We can assume that some $P(t)$ is not the null matrix, or there would be nothing to prove. The matrices $P(t)$ form a family $\mathfrak{M}$ satisfying the conditions of Theorem 3. Moreover each $p_{ij}(t)$ is continuous, if $t>0$, by Theorem 1, and so for all $t$, from (0.1). Using the notation of Theorem 3, if $i \notin F$, $p_{ij}(t)=u_j$ or $p_{ij}(t)=0$. Then if $i \notin F$, $p_{ij}(t)$ is independent of $t$. This means that $\mathfrak{M}'$ consists only of the identity, so $P(t)$ is independent of $t: P(t) \equiv U$. The example above shows that the measurability of the $p_{ij}(t)$ is a necessary part of the hypotheses.

Theorem 4. *If the* $p_{ij}(t)$ *satisfying* (0.1) *are continuous, then* $\lim_{t \to 0} P(t) = U$ *exists. The matrix U is a non-null matrix of the type described in Theorem* 2, *and* (9)

$$UP(t) \leqq P(t)U = P(t). \tag{4.1}$$

*(In the following we shall use the notation of Theorem* 2.) *Moreover*

$$p_{ij}(t) = 0 \qquad (j \in F). \tag{4.2}$$

*There are continuous functions* $\Pi_{IJ}(t)$, *satisfying* (0.1) and

(9) An inequality between two matrices is defined to mean the same inequality between their corresponding elements.

$$\lim_{t\to 0} \Pi_{IJ}(t) = \delta_{IJ} \tag{4.3}$$

*such that*

$$p_{ij}(t) = \Pi_{IJ}(t)u_j \qquad (i \in I, j \in J). \tag{4.4}$$

*There are continuous functions* $\Pi_{iJ}(t)$ $(i\in F)$ *such that*(10)

$$\begin{gathered} p_{ij}(t) = \Pi_{iJ}(t)u_j \qquad (i \in F, j \notin F), \\ \Pi_{iK}(t) \geqq \sum_J p_{iJ}\, \Pi_{JK}(t). \end{gathered} \tag{4.5}$$

*Conversely, if the* $p_{ij}(t)$ *satisfy* (0.1) *and if* $\lim_{t\to 0} P(t)$ *exists, the* $p_{ij}(t)$ *are continuous.*

Neglecting subscripts in $F$, this theorem reduces the study of $P(t)$ to that of $(\Pi_{IJ}(t))$ in which case the limit matrix $(t\to 0)$ is the identity.

Let $U:(u_{ij})$, $U':(u_{ij}')$ be limiting matrices of $P(t)$, $t\to 0$. Then (0.1) implies (4.1). The equal $i$th row sums in (4.1) are

$$\sum_j p_{ij}(t) \sum_k u_{jk} = \sum_k p_{ik}(t) = 1. \tag{4.6}$$

Since the row sums of $U$ are less then or equal to 1, (4.6) implies that if $\sum_k u_{jk} < 1$, $p_{ij}(t) = 0$. Then in this case $u_{ij} = u_{ij}' = 0$ also. It follows from (4.1) that

$$\sum_j u_{ij}' u_{jk} \leqq u_{ik}' \qquad (U'U \leqq U'). \tag{4.7}$$

Summing over $k$, since $u_{ij}'=0$ if $\sum_k u_{jk}<1$, we see that both sides of (4.7) have sum $\sum_k u_{ik}'$; so there is equality in (4.7):

$$U'U = U'. \tag{4.7'}$$

Replacing $U$ by $U'$ in the inequality $UP(t) \leqq P(t)$, and letting $t$ approach 0 in such a way that $P(t)\to U$, we obtain

$$U'U \leqq U. \tag{4.8}$$

Then combining (4.7′) and (4.8), we have $U' \leqq U$, and by symmetry $U \leqq U'$; so $U = U'$. There is thus only one limiting matrix $U: P(t)\to U$. Since equation (4.7′) becomes $U^2 = U$, Theorem 2 is applicable. In the following, we shall use the notation of that theorem. If $k\in F$, $u_{jk}=0$; therefore (using (4.1)) $p_{ik}(t)=0$ also, for all $i$. Then $U$ is not the null matrix. If $i, k \notin F$, (4.1) implies

$$\left(\sum_{j\in K} p_{ij}(t)\right) u_k = p_{ik}(t) \qquad (k \in K), \tag{4.9}$$

(10) If the $G_\nu$ contain only one subscript each, so that $p_{ij}(t)\to\delta_{ij}$ $(t\to 0)$ if $i \notin F$, then we can read $p_{ij}(t)$ for $\Pi_{IJ}(t)$, $p_{ij}(t)$ for $\Pi_{IJ}(t)$ throughout.

and

$$\sum_{j \in I} u_j p_{jk}(t) \leq p_{ik}(t) \qquad (i \in I). \tag{4.10}$$

There is equality in (4.10) (and we shall refer to it as if it were so written), because summing over $k$ gives 1 on both sides. Equations (4.9) and (4.10) imply that if $i, k \notin F$, $p_{ik}(t)/u_k$ depends only on $I, K$: $p_{ik}(t) = \Pi_{IK}(t)u_k$. Evidently the matrix function $(\Pi_{IJ}(t))$ satisfies (0.1), and (4.3) is true. If $i \in F$, and $k \notin F$, (4.1) implies

$$\sum_J p_{iJ}\,\Pi_{JK}(t) \leq \left(\sum_{j \in K} p_{ij}(t)\right) u_k = p_{ik}(t), \tag{4.11}$$

so that if $\Pi_{iK}(t)$ is defined as the parentheses in (4.11), (4.5) follows at once.

Conversely, if $P(t) \to U$ $(t \to 0)$,

$$\lim_{t \downarrow 0} P(s+t) = P(S)U. \tag{4.12}$$

The function $p_{ij}(t)$ having a right-hand limit for all $t$ has at most denumerably many discontinuities, is therefore measurable, and continuous (Theorem 1).

THEOREM 5. *Let $\alpha$ be a given subscript. Then if $P(t)$ satisfies* (0.1), *$p_{\alpha j}(t)$ will be continuous and $\lim_{t \to 0} p_{\alpha j}(t)$ will exist, for all $j$, if $\sum_j p_{\alpha j}(t)$ converges uniformly in some interval $0 < t < t_0$.*

Doeblin (I) proved that if $P(t)$ satisfies (0.1) and is finite-dimensional, then the $p_{ij}(t)$ are continuous and have unique limits as $t \to 0$. This fact which evidently is a consequence of Theorem 5, can be proved directly as follows. Let $\mathfrak{M}$ be the set of limiting matrices of $P(t)$, $t \to 0$. Then $\mathfrak{M}$ satisfies the conditions of Theorem 3, Corollary 1, so $\mathfrak{M}$ contains only a single matrix $U$. It follows that $P(t) \to U$, and the $p_{ij}(t)$ are then continuous, by Theorem 4.

**Proof of Theorem** 5. Let $G$ be the set of subscripts $\alpha$ with the property described in the theorem. The equation $P(s)P(t) = P(t)P(s)$ implies that if $A : (a_{ij})$ is a limiting matrix of $P(t)$, $t \to 0$, then

$$\sum_j a_{ij} p_{jk}(t) \leq \sum_j p_{ij}(t) a_{jk}. \tag{5.1}$$

If $i \in G$, then $\sum_j a_{ij} = 1$, and the sum over $k$ on the left is 1, so that on the right is also 1. Then there is equality in (5.1):

$$\sum_j a_{ij} p_{jk}(t) = \sum_j p_{ij}(t) a_{jk} \qquad (i \in G). \tag{5.1'}$$

If $\sum_k a_{jk} < 1$, then $p_{ij}(t) = 0$, or the sum over $k$ on the right in (5.1′) would not be 1. If $j \notin G$, we can find an $A$ with $\sum_k a_{jk} < 1$, whence it follows that

$p_{ij}(t)=0$ if $i\in G$, $j\notin G$. Let $P'(t)$ be the matrix obtained by dropping all elements of $P(t)$ with a subscript not in $G$. Then $P'(t)$ satisfies (0.1) and has the property that any limiting matrix ($t\to 0$) has row sums 1. The proof given above of Doeblin's result goes through word for word, applied to $P'(t)$. We have thus proved that $P_{ij}(t)$ is continuous, and $\lim_{t\to 0} p_{ij}(t)$ exists, if $\alpha\in G$, and in addition that $p_{ij}(t)\equiv 0$ if $\alpha\in G, j\notin G$.

We now turn to an examination of the limiting matrices of $P(t)$, as $t\to\infty$.

THEOREM 6. *Define the matrix* $U:(u_{ij})$ *by*

$$\liminf_{t\to\infty} p_{ij}(t) = u_{ij}. \tag{6.1}$$

*Then*

(a) *$U$ is a limiting matrix of $P(t)$, as $t\to\infty$; $U$ has the properties described in Theorem* 2, *and* $P(t)U=UP(t)=U$;

(b) (6.1) *can be sharpened to*

$$\lim_{t\to\infty} p_{ij}(t) = u_{ij} \tag{6.1'}$$

*if $i$ is a subscript such that* $\sum_j u_{ij}=1$[11].

(c) *Using the notation of Theorem* 2, *and assuming that $U$ is not the null matrix,*

$$\begin{aligned} p_{ij}(t) &= 0 && (i\in I, j\notin I),\\ \sum_{r\in J} u_r p_{rj}(t) &= u_j && (j\in J),\\ \sum_{j\in K} p_{ij}(t) + \sum_{j\in F} p_{ij}(t)\rho_{jK} &= \rho_{iK} && (i\in F). \end{aligned} \tag{6.2}$$

*If $i\notin F$ or if $j\notin F$* (6.1′) *is true. If $i\notin F$, $p_{ij}(t)$ is continuous, and* $\lim_{t\to 0} p_{ij}(t)$ *exists. Moreover*

$$\lim_{t\to\infty} \sum_{r\in K} p_{ir}(t) = \rho_{iK}, \qquad \lim_{t\to\infty} \sum_{j\in F} p_{ij}(t)\rho_{jK} = 0 \qquad (i\in F). \tag{6.3}$$

The fact that if $P(t)$ is finite-dimensional (6.1′) is always true, which follows from Theorem 6, can be proved directly as follows. The set of limiting matrices (in this case) of $P(t)$, $t\to\infty$, is seen at once to have the properties required in Theorem 3, Corollary 1, so there is only one limiting matrix $U$: $P(t)\to U$. This argument breaks down in the infinite-dimensional case, in which a more detailed analysis is necessary.

Let $\mathfrak{L}$ be the class of limiting matrices $(a_{ij})$ of $P(t)$, as $t\to\infty$. Then $\mathfrak{L}$ includes all its limit matrices. This implies that $\mathfrak{L}$ contains one or more matrices minimizing $\sum_{i,j}2^{-i}a_{ij}$. Let $\mathfrak{M}$ be the class of these minimal matrices. We shall show that $\mathfrak{M}$ contains only one matrix, $U$, defined by (6.1). The proof will be carried through in several steps.

[11] It follows that if the matrices are finite-dimensional, (6.1′) is true for all $i$, $j$, a fact due to Doeblin (1).

(i) *If* $A\in\mathfrak{L}$, *then* $P(t)A\in\mathfrak{L}$.

This can be deduced at once from (0.1).

(ii) *If* $A\in\mathfrak{M}$, *then* $AP(t)=P(t)A\in\mathfrak{M}$.

This follows, for (0.1) implies that if $A:(a_{ij})\in\mathfrak{M}$, there is a $B:(b_{ij})\in\mathfrak{L}$, depending on $t$ and on $A$, such that

$$\sum_j b_{ij}p_{jk}(t) \leqq \sum_j p_{ij}(t)b_{jk} = a_{ik}. \tag{6.4}$$

Summed over $k$, this means that

$$\sum_j b_{ij} \leqq \sum_k a_{ik} \qquad (i = 1, 2, \cdots) \tag{6.5}$$

and only equality is possible, since $A\in\mathfrak{M}$. Then (6.4) becomes the equality $BP(t)=P(t)B=A$. Therefore,

$$AP(t) = P(t)BP(t) = P(t)A.$$

By (i), $P(t)A\in\mathfrak{L}$, and summing over the $i$th row of $P(t)A=AP(t)$ gives $\sum_j a_{ij}$, so $P(t)A\in\mathfrak{M}$, since $A\in\mathfrak{M}$.

(iii) *If* $A, B\in\mathfrak{M}$, *then* $AB=BA\in\mathfrak{M}$.

By (ii), if $A\in\mathfrak{M}$, it follows that $AP(t)\in\mathfrak{M}$. If $B\in\mathfrak{M}$, $AB$ is a limiting matrix of $AP(t)$, $t\to\infty$, so $AB\in\mathfrak{M}$ since $\mathfrak{M}$ is closed. Moreover by (ii), $AP(t)=P(t)A$, so $AB\geqq BA$. By symmetry, the reverse inequality is also true; so $AB=BA$.

(iv) *If* $A\in\mathfrak{L}$ *there is an* $A'\in\mathfrak{M}$ *with* $A'\leqq A$.

For if $A\in\mathfrak{L}$, there is, using (0.1), a $B\in\mathfrak{M}$ and a $C\in\mathfrak{L}$ such that $A\geqq BC$, and since $BC\in\mathfrak{M}$ (as a limit of $BP(t)\in\mathfrak{M}$) this is the desired inequality.

(v) *If* $A, B\in\mathfrak{M}$, *there is a* $C\in\mathfrak{M}$ *with* $A=BC$.

We see this, for there is certainly, using (0.1), a $C\in\mathfrak{L}$ with $A\geqq BC$. As we have seen, $BC\in\mathfrak{M}$, so there must be equality.

(vi) *If* $A\in\mathfrak{M}$, *and if n is any positive integer, there is a* $B\in\mathfrak{M}$ *such that* $A=B^n$.

For, since $P(t/n)^n=P(t)$, if $A\in\mathfrak{M}$, there is a $B_1\in\mathfrak{L}$ such that $B_1^n\leqq A$. By (iv) there is then a $B\in\mathfrak{M}$ with $B^n\leqq B_1^n\leqq A$. To show that there must be equality, we need only show that $B^n\in\mathfrak{L}$. Since $B\in\mathfrak{M}$, $BB=B^2\in\mathfrak{M}$ by (iii). Then $BB^2=B^3\in\mathfrak{M}$, and so on.

Now (iii) and (v) imply that the matrices of $\mathfrak{M}$ form a commutative group The fact that $\mathfrak{M}$ is closed and that (vi) is true shows that $\mathfrak{M}$ has the properties required in Theorem 3, Corollary 1. There can therefore be only a single matrix $U:(u_{ij})$ in $\mathfrak{M}$, and $U$ has the properties described in Theorem 2. Because of (iv), $\liminf_{t\to\infty} p_{ij}(t)=u_{ij}$. From now on we shall assume the notation of Theorem 2. The equality $P(t)U=UP(t)=U$ follows from (ii). If $\sum_j u_{\alpha j}=1$, no limiting value of $p_{\alpha j}(t)$, $t\to\infty$, can be greater than $u_{\alpha j}$, or there would be a limiting row having a sum greater than 1. Then if $\sum_j u_{\alpha j}=1$, $\lim_{t\to\infty} p_{\alpha j}(t)=u_{\alpha j}$, for all $j$. In particular, (6.1′) is true if $i\notin F$. The equations of (6.2) are equiva-

lent to the equations $P(t)U=UP(t)=U$. If $(a_{ij})$ is a limiting matrix of $P(t)$ as $t\to 0$, (6.2) implies that

$$\sum_{r\in J} u_r a_{rj} = u_j \qquad (j \in J). \tag{6.6}$$

Summing (6.6) over $j\in J$ we obtain

$$\sum_{r\in J} u_r\Big(\sum_{j\in J} a_{rj}\Big) = \sum_{j\in J} u_j = 1. \tag{6.7}$$

Then $\sum_j a_{rj}=1$ if $r\in J$. This implies that $\sum_j p_{rj}(t)$ converges uniformly in some interval $0<t<t_0$; so according to Theorem 5, $p_{rj}(t)$ is continuous, and has a unique limit as $t\to 0$, if $r\in J$. Then this is true for any subscript $r\notin F$. As $t\to\infty$ in the last equation of (6.2) the first sum has an inferior limit greater than or equal to $\rho_{iK}$. Then there must actually be convergence; the first equation of (6.3) is true. The second sum in the last equation of (6.2) must then approach 0; (6.3) is true. Equation (6.3) is impossible, since lim $\inf_{t\to\infty} p_{ik}(t)=\rho_{iK}u_k$, $(i\in F,\ k\in K)$ unless $p_{ik}(t)\to\rho_{iK}u_k$; so (6.1′) is true if $j\notin F$. The proof of the theorem is now complete.

Regularity hypotheses imposed on the probability matrices can be used to simplify the above results. Thus suppose that there is a value $t_0$ of $t$ such that $\sum_j p_{ij}(t_0)$ converges uniformly in $i$. It follows readily that $\sum_j p_{ij}(t)$ converges uniformly in $i$ and $t\geqq t_0$. This means that any limiting matrix of $P(t), t\to\infty$, has row sums 1, so $P(t)\to U$, by Theorem 6. A less strong condition is that there be a value $t_0$ of $t$, a positive integer $N$ and a positive $\epsilon$ such that $\sum_{j\leqq N} p_{ij}(t_0)\geqq\epsilon$ for all $i$. It follows readily that the same inequalities hold for $t\geqq t_0$. Then

$$\sum_{k\leqq N} u_{jk} \geqq \epsilon; \tag{6.8}$$

so there can only be a finite number of $G_\nu$, and $U$ cannot be the null matrix. Also if $j\in F$, (6.8) becomes

$$\sum_{k\leqq N} \rho_{jK} u_k \geqq \epsilon \qquad (k \in K). \tag{6.8′}$$

Then some $\rho_{jK}>0$ for each $j\in F$, so by (6.3), $\lim_{t\to\infty} p_{ij}(t)=0$, if $i\in F$, $j\in F$. Thus in this case also, $P(t)\to U$, as $t\to\infty$. The fact that $P(t)\to U$ under the above hypotheses can also be derived using general theorems of Doeblin(12) or of Kryloff and Bogolioùboff(13).

If there is a set of non-negative numbers $p_1, p_2, \cdots$ such that

$$\sum_i p_i p_{ij}(t) = p_j \quad (\text{all } j), \qquad \sum_j p_j = 1, \tag{6.9}$$

(12) Thesis, Paris, 1938, pp. 105–109.

(13) Paris, Comptes Rendus de l'Académie des Sciences, vol. 204 (1937), pp. 1454–1456.

the set $p_1, \cdots$ will be called a set of (stationary) absolute probabilities. The number $p_j$ can be considered as the probability of being in the $j$th state at time $t$. Any linear combination of absolute probabilities with non-negative coefficients is also a set of absolute probabilities, or proportional to a set. If $U$ is defined as in Theorem 6, the second set of equations of (6.2) states that the $i$th row of $U$, if $i \notin F$, is a set of absolute probabilities. If $i \in F$, the $i$th row of $U$ is a linear combination (coefficients $\rho_{iK}$) of the rows of elements with first subscripts not in $F$. Then every row of $U$ is a set of absolute probabilities, or proportional to a set (if the row sum is less than 1). Moreover (6.9) implies that $\sum_i p_i u_{ij} = p_j$; so any set of absolute probabilities is a linear combination (non-negative coefficients) of rows of $U$. The states with subscripts in $F$ then always have probability 0, regardless of the absolute probabilities. One simple consequence of these remarks is that if there is a solution to (6.9), $U$ cannot be identically 0, and some row of $U$ is also a solution of (6.9); there is a solution of (6.9) determined by the equations $p_j = \lim_{t\to\infty} p_{\alpha j}(t)$, $\alpha$ fixed, not in $F$.

THEOREM 7. *Suppose that the $p_{ij}(t)$ satisfying* (0.1) *are continuous. Then if $U$ is defined as in Theorem* 6,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T p_{ij}(t)dt = u_{ij} \tag{7.1}$$

*for all $i, j$.*

Let $Q(T)$ be the matrix with general element $q_{ij}(T)$:

$$q_{ij}(T) = \frac{1}{T}\int_0^T p_{ij}(t)dt.$$

Since

$$\sum_j q_{ij}(T)p_{jk}(t) = \sum_j p_{ij}(t)q_{jk}(T) = \frac{1}{T}\int_t^{T+t} p_{ik}(s)ds, \tag{7.2}$$

if $U'$ is a limiting matrix of $Q(T)$, $T\to\infty$, it follows that

$$U'P(t) \leqq P(t)U' = U'.$$

Since the row sums of $U'P(t)$ are the same as those of $U'$, there must be equality:

$$U'P(t) = P(t)U' = U'. \tag{7.3}$$

According to Theorem 6,

$$UP(t) = P(t)U = U. \tag{7.4}$$

It follows from (7.3) and (7.4) that

$$UU' \leqq U'U = U', \quad U'U \leqq UU' = U. \tag{7.5}$$

Then $U = UU' \leqq U' = U'U \leqq U$, $U = U'$, as was to be proved.

The following is a simple example illustrating the fact that $U$ in Theorems 6, 7 may be the null matrix. Let $p_{ij}(t) = 0$ if $j < i$, and otherwise define $p_{ij}(t)$ by

$$p_{ij}(t) = \frac{t^{j-i}}{(j-i)!} e^{-t}. \tag{7.6}$$

Evidently $p_{ij}(t) \to 0$, as $t \to \infty$. There can be no stationary absolute probabilities in this case.

In examining the successive transitions of the system, we shall assume that the system is initially in a state $\alpha$, where $\alpha$ will be held fixed throughout the discussion. Let $\xi(t)$ be the number of the state assumed by the system at time $t$. Then $\xi(t)$, for each fixed value of $t$, is a chance variable: $\xi(0) \equiv \alpha$; $\xi(t) = j$ with probability $p_{\alpha j}(t)$ if $t > 0$. To discuss the continuity properties of $\xi(t)$ in $t$ we shall assume a minimum of regularity properties of $P(t)$, to which we shall be led in a natural way. In order to discuss the probability measures under consideration, we must, as usual, find a space $\Omega^*$ of points $\omega$, a measure defined on $\Omega^*$, and a one-parameter family of measurable functions $x_t(\omega)$, $0 \leqq t < \infty$, such that the probability relations of the chance variables $\{\xi(t)\}$ become measure relations of the functions $\{x_t(\omega)\}$. Let $\Omega^*$ be the space of all functions $x(t)$, $0 \leqq t < \infty$, taking on the integral values used in the subscripts of $P(t)$. A probability measure on $\Omega^*$ is defined as follows. If $0 = t_0 < t_1 < \cdots < t_n$, the conditions

$$x(t_j) = \nu_j \qquad (j = 0, \cdots, n) \tag{8.1}$$

determine a subset of $\Omega^*$ and the measure of this subset is defined by

$$\begin{aligned} P^*\{x(t_j) = \nu_j, j = 0, \cdots, n\} \\ = \delta_{\alpha\nu_0} p_{\alpha\nu_1}(t_1) p_{\nu_1\nu_2}(t_2 - t_1) \cdots p_{\nu_{n-1}\nu_n}(t_n - t_{n-1}). \end{aligned} \tag{8.2}$$

By a theorem of Kolmogoroff[14], a completely additive measure function is determined on $\Omega^*$ by these conditions. Let $x_s(\omega)$ be the function of $\omega: x(t)$ which takes on the numerical value $f(s)$ if $\omega$ is the function $f(t)$. Then the probability relations of the chance variables $\{\xi(t)\}$ become measure relations of the measurable functions $\{x_t(\omega)\}$:

$$P^*\{x_t(\omega) = j\} = p_{\alpha j}(t)$$

and so forth. We shall sometimes write $x(t)$ instead of $x_t(\omega)$, so that "$x(t)$" can mean, for example: (a) a point $\omega$ of $\Omega^*$; (b) a function $x_t(\omega)$ of $\omega$; (c) a num-

[14] *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 4, pp. 24–30. The fact that our functions assume only integral values, whereas those of Kolmogoroff assume all values necessitates only trivial changes in the proof.

ber, the value of the function $x(t)$ at the point $t$. When there is any danger of confusion, the proper meaning will be explicitly stated. The function $x_t(\omega)$ is automatically defined on any subset of $\Omega^*$, and it is usually desirable to restrict $\omega$ to be in a subset $\Omega$ of $\Omega^*$, of outer measure 1, defining a $P$-measure on $\Omega$ by setting $P(\Lambda^*\cdot\Omega)=P^*(\Lambda^*)$ for any $P^*$-measurable set[15] $\Lambda^*$. The probability relations of the chance variables $\{\xi(t)\}$ now become the measure relations of the functions $x_t(\omega)$, $\omega\in\Omega:P\{x_s(\omega)=j\}=p_{\alpha j}(s)$ and so on[16]. It has been shown[17] that if any $P^*$-measure is given, there always corresponds an everywhere dense denumerable sequence of real numbers $R:\{r_j\}$ such that if $I$ is any open interval, and if $s\in I$,

$$(8.3)\qquad P^*\left\{\underset{r_j\in I}{\text{G.L.B.}}\ x(r_j) \leqq x(s) \leqq \underset{r_j\in I}{\text{L.U.B.}}\ x(r_j)\right\} = 1^{(18)}.$$

It has been shown[19] that $\Omega$ can be chosen to consist of all (possibly infinite-valued) functions $x(t)$ which satisfy the relation

$$(8.4)\qquad \liminf_{r_j\to t} x(r_j) \leqq x(t) \leqq \limsup_{r_j\to t} x(r_j)$$

for all $t\in R$. Then if this is done,

$$(8.5)\qquad \underset{r_j\in I}{\text{G.L.B.}}\ x(r_j) \leqq x(s) \leqq \underset{r_j\in I}{\text{L.U.B.}}\ x(r_j)$$

for all $s\in I$, $\omega:x(t)$ in $\Omega$, sharpening (8.3). Such a space $\Omega$ is called quasi-separable, and the process: that is, the combination of $\Omega$ with its $P$-measure, is called a quasi-separable process.

A measure can be defined on the space $T\times\Omega$ of couples $(t,\omega)$, as the product of Lebesgue measure on the $t$-axis and $P$-measure on $\omega$. The process is called measurable if the function $x_t(\omega)$ is $(t,\omega)$-measurable. The $P^*$-measure is then said to determine a measurable process. This hypothesis on the $P^*$-measure is certainly a minimum hypothesis. On the other hand, there are natural analytic restrictions on the $p_{ij}(t)$. Let $G_\alpha$ be the set of subscripts $j$ such that $p_{\alpha j}(t)\not\equiv 0$. Only the subscripts in $G_\alpha$ need be considered in analyzing the transitions of the system, supposed initially in state $\alpha$. It follows readily from (0.1) that $p_{ij}(t)\equiv 0$ if $i\in G_\alpha$, $j\notin G_\alpha$. The matrix $P_\alpha(t):p_{ij}(t)$ with $i,j\in G_\alpha$ then satisfies (0.1), and it is this matrix $P_\alpha(t)$ which is essential to the discussion. The

(15) Cf. Doob, these Transactions, vol. 42 (1937), pp. 108–110.

(16) $P\{x_s(\omega)=j\}$ is to be interpreted as the $\Omega$-measure of the set of all functions $x(t)$ in $\Omega$ for which $x(s)=j$.

(17) Doob, these Transactions, vol. 47 (1940), p. 467.

(18) This equality holds for each fixed $s$. The $\omega$-set $\{x(s)\leqq\phi,\ s\in I\}$, $\phi$ a $P^*$-measurable function, is not $P^*$-measurable, so each value of $s$ must be considered separately in (8.3), or in probability relations of similar type. The subspace $\Omega$ is introduced in order to avoid this necessity.

(19) Op. cit., pp. 468–469.

natural analytic hypotheses on $P_\alpha(t)$ would include the measurability of its elements. This, by Theorem 1, implies their continuity, and then (Theorem 4), $\lim_{t\to 0} P_\alpha(t) = U$ exists. The matrix $U$ is the first determining factor of the regularity of the process. It is natural to suppose that it is the identity matrix. A glance at Theorem 4 shows that no other hypothesis can possibly be compatible with any sort of continuity in the transitions of the system.

These considerations lead to the following formulation of a natural hypothesis to be imposed on the matrix function $P_\alpha(t)$. We shall denote as hypothesis $H_\alpha$ the hypothesis that the system is initially in state $\alpha$, and that if $i \in G_\alpha$, $\lim_{t\to 0} p_{ii}(t) = 1$. Then $\lim_{t\to 0} p_{ij}(t) = \delta_{ij}$ $(i \in G_\alpha)$. Since $P_\alpha(t)$ satisfies (0.1), the $p_{ij}(t)$ $(i, j \in G_\alpha)$ will be continuous (Theorem 4). Moreover $p_{ij}(t) \equiv 0$ if $i \in G_\alpha, j \notin G_\alpha$. Then if hypothesis $H_\alpha$ is true, and if $i \in G_\alpha$, $p_{ij}(t)$ is continuous for all $j$, and $\lim_{t\to 0} p_{ij}(t) = \delta_{ij}$. Hypothesis $H_\alpha$ implies the continuity of $p_{\alpha j}(t)$ for all $j$, even though $\alpha$ may not be in $G_\alpha$. In fact the equation

$$p_{\alpha j}(s + h) = \sum_{i \in G_\alpha} p_{\alpha i}(s) p_{ij}(h)$$

shows that $p_{\alpha j}(t)$ is continuous for $t > s$, and therefore for all $t$. If $i \in G_\alpha$, then $p_{ii}(t) > 0$ for all $t$, and if $i = \alpha$, $j \in G_\alpha$ or if $i, j \in G_\alpha$ then $p_{ij}(t) = 0$ at most on a finite interval $0 < t \leqq t_0$ (depending on $i, j$). The first fact follows from the inequality $p_{ii}(t) \geqq p_{ii}(t/n)^n$, $n = 1, 2, \cdots$, since $\lim_{n\to\infty} p_{ii}(t/n) = 1$. The second fact follows from the inequality $p_{ij}(t+h) \geqq p_{ij}(t) p_{jj}(h)$ which implies that if $p_{ij}(t') > 0$, then $p_{ij}(t) > 0$, for $t > t'$.

The following theorem shows the relations between various hypotheses it would be natural to assume.

THEOREM 8. *Suppose that* $P^*\{x(0) = \alpha\} = 1$. *Then the following three conditions on* $P^*$*-measure are equivalent.*

(i) *The* $P^*$*-measure determines a measurable process.*

(ii) *Hypothesis* $H_\alpha$ *is satisfied.*

(iii) *For every* $\tau > 0$,

$$\lim_{t\to\tau} P^*\{x(t) = x(\tau)\} = 1. \tag{8.6}$$

In the usual language of measure theory, (8.6) states that $x_t(\omega) \to x_\tau(\omega)$ in measure. We shall prove a much stronger result below, Theorem 11. To prove Theorem 8, we prove that (i) implies (ii), (ii) implies (iii), and (iii) implies (i).

**Proof that (i) implies (ii).** Suppose that $P^*$-measure determines a measurable process. Then it follows([20]) that for fixed $h > 0$, $\epsilon > 0$, $P^*\{|x(t+h) - x(t)| > \epsilon\}$ is Lebesgue measurable in $t$, and (as $h \to 0$) goes to 0 in measure on any finite $t$-interval. If $\epsilon < 1$,

$$P^*\{|x(t + h) - x(t)| > \epsilon\} = \sum_{j \in G_\alpha} p_{\alpha j}(t)[1 - p_{jj}(h)]. \tag{8.7}$$

([20]) Doob, these Transactions, vol. 42 (1937), p. 117.

Since the quantity in (8.7) goes to 0 as $h\to 0$ in measure on every finite $t$-interval, and since $p_{\alpha j}(t)>0$ if $t$ is sufficiently large, $\lim_{h\to 0} p_{jj}(h)=1$ if $j\in G_\alpha$. Then hypothesis $H_\alpha$ is satisfied.

**Proof that** (ii) **implies** (iii). If hypothesis $H_\alpha$ is true, we shall prove (8.6) by evaluating the probabilities involved. If $0<\tau<t$,

$$(8.8)\qquad \begin{aligned} P^*\{x(t)=x(\tau)\} &= \sum_j P^*\{x(t)=x(\tau)=j\} \\ &= \sum_{j\in G_\alpha} p_{\alpha j}(\tau)p_{jj}(t-\tau) \end{aligned}$$

and if $0<t<\tau$,

$$(8.8')\qquad P^*\{x(t)=x(\tau)\} = \sum_{j\in G_\alpha} p_{\alpha j}(t)p_{jj}(\tau-t).$$

**Proof that** (iii) **implies** (i). Condition (iii) is known (loc. cit. ([20])) to imply that the $P^*$-measure determines a measurable process.

Now the series in (8.8) is majorized by $\sum_j p_{\alpha j}(\tau)$, and that in (8.8′) by $\sum_j p_{\alpha j}(t)$. Then the series in (8.8) converges uniformly in $t$. The series $\sum_j p_{\alpha j}(t)$ is a series of non-negative continuous functions, converging to the continuous function 1, so there is uniform convergence in a neighborhood of $\tau$. Thus the series in (8.8) and (8.8′) are uniformly convergent for $t$ near $\tau$, and when $t\to\tau$ both become $\sum_j p_{\alpha j}(\tau)=1$, as was to be proved.

THEOREM 9. *Suppose that hypothesis $H_\alpha$ is true. Then if $i\in G_\alpha$,*

$$(9.1)\qquad \lim_{t\to 0}\frac{1-p_{ii}(t)}{t} = q_i \ (\leqq +\infty)$$

*exists. If $q_i=0$, $p_{ii}(t)\equiv 1$. If $q_i=\infty$, then if $j\neq i$, $j\in G_\alpha$,*

$$(9.2)\qquad \lim_{t\to 0}\frac{p_{ij}(t)}{1-p_{ii}(t)} = \lim_{t\to 0}\frac{p_{ji}(t)}{1-p_{ii}(t)} = 0.$$

*If $q_i<\infty$, then for $j\neq i$, $j\in G_\alpha$ the limits*

$$(9.3)\qquad \lim_{t\to 0}\frac{p_{ij}(t)}{t} = q_{ij}\ (<+\infty),\qquad \lim_{t\to 0}\frac{p_{ji}(t)}{t} = q_{ji}\ (<+\infty)$$

*exist, and*

$$(9.4)\qquad \sum_j q_{ij} \leqq q_i.$$

*In the finite-dimensional case*([21]), *$q_i<\infty$ for all $i\in G_\alpha$, and there is equality in* (9.4).

Let $R$ be a denumerable everywhere dense $t$-set. Since when $t\to\tau$,

([21]) Doeblin (I) proved Theorem 9 in the finite-dimensional case.

$x(t) \to x(\tau)$ in measure (Theorem 8), it follows that

$$\liminf_{r\to\tau} x(r) \leqq x(\tau) \leqq \limsup_{r\to\tau} x(r) \qquad (r \in R) \tag{9.5}$$

with probability 1 (that is, almost everywhere on $\Omega^*$). Then (8.3) is satisfied. We shall also need the following fact: If $I$ is any open $t$-interval, and if $t_1^{(n)} < \cdots < t_{\nu_n}^{(n)}$ are points in $I$, with $\max_j (t_j^{(n)} - t_{j-1}^{(n)}) = \delta_n$, then if $\delta_n \to 0$,

$$\lim_{n\to\infty} \text{L.U.B.}_{j}\, x(t_j^{(n)}) = \text{L.U.B.}_{r\in R\cdot I}\, x(r) \tag{9.6}$$

with probability 1. This can be proved as follows. Because of the fact that when $t\to\tau$, $x(t)\to x(\tau)$ in measure, it surely is true that for each $r$ in $I$,

$$\liminf_{n\to\infty} \text{L.U.B.}_{j}\, x(t_j^{(n)}) \geqq x(r) \tag{9.7}$$

with probability 1, and (9.7) implies (9.6), because of (8.3). In the same way we can prove

$$\lim_{n\to\infty} \text{G.L.B.}_{j}\, x(t_j^{(n)}) = \text{G.L.B.}_{r\in R\cdot I}\, x(r) \tag{9.8}$$

with probability 1.

Now let $i \in G_\alpha$, and choose $\tau$ so that $p_{\alpha i}(\tau) > 0$. Let $\phi_i(h)$ be the probability that if $x(\tau) = i$ then $x(r) = i$ for $\tau \leqq r \leqq \tau + h$ $(r \in R)$. (If $R$ is used to determine a quasi-separable process, $\phi_i(h)$ is the probability that if $x(\tau) = i$, then $x(t) = i$ for $\tau \leqq t \leqq \tau + h$.) According to (9.6) and (9.8), if $\tau = t_1^{(n)} < \cdots \leqq h$, and $\max_j (t_j^{(n)} - t_{j-1}^{(n)}) = \delta_n \to 0$, then

$$\begin{aligned} P^*\{x(r) = i, \tau \leqq r \leqq \tau + h, (r \in R)\} &= \lim_{n\to\infty} P^*\{x(t_j^{(n)}) = i, j \geqq 1\} \\ &= p_{\alpha i}(\tau) \lim_{n\to\infty} \prod_{j\geqq 1} p_{ii}(t_{j+1}^{(n)} - t_j^{(n)}) = p_{\alpha i}(\tau)\phi_i(h). \end{aligned} \tag{9.9}$$

Let $\{\epsilon_n\}$ be any sequence of positive numbers converging to 0. To prove (9.1) it will be sufficient to let $t \to 0$ through the sequence $\{\epsilon_n\}$, and to show that there is a limit, which is independent of the sequence $\{\epsilon_n\}$. Choose the integers $m_n$ so that $m_n\epsilon_n \uparrow h$. Then setting $t_{j+1}^{(n)} - \tau = j\epsilon_n$ in (9.9), $0 \leqq j \leqq m_n$,

$$\lim_{n\to\infty} p_{ii}(\epsilon_n)^{m_n} = \phi_i(h). \tag{9.10}$$

This implies that if $\phi_i(h) > 0$

$$\lim_{n\to\infty} m_n \log p_{ii}(\epsilon_n) = -\lim_{n\to\infty} h\,\frac{1 - p_{ii}(\epsilon_n)}{\epsilon_n} = \log \phi_i(h). \tag{9.11}$$

We have thus shown that unless $\phi_i(h) = 0$, (9.1) is true, and

$$(9.12) \qquad \phi_i(h) = e^{-q_i h}.$$

On the other hand, if $\phi_i(h)=0$, (9.1) is true with $q_i=\infty$, and then $\phi_i(h)\equiv 0$. Since $p_{ii}(t)\geqq\phi_i(t)$, $q_i=0$ implies that $p_{ii}(t)\equiv 1$. In proving (9.2) and (9.3) we can assume that $p_{ii}(t)<1$ for all $t$, since otherwise (0.1) implies that $p_{ii}(t)\equiv 1$, so $q_i=0$: in this case (9.2) is inapplicable; the first part of (9.3) is obvious, and the second is proved by a trivial modification of the proof below. To prove (9.2) we note that if $\eta>0, j\neq i, j\in G_\alpha$,

$$(9.13) \qquad p_{ij}(n\epsilon) \geqq \sum_{k=0}^{n-1} p_{ii}(\epsilon)^k p_{ij}(\epsilon) p_{jj}(\overline{n-k-1}\,\epsilon) \geqq (1-\eta)\frac{1-p_{ii}(\epsilon)^n}{1-p_{ii}(\epsilon)} p_{ij}(\epsilon),$$

if $n\epsilon$ is sufficiently small. Then if $q_i=\infty$, when $n\to\infty$ and $\epsilon\to 0$ so that $n\epsilon\to t$, (9.13) becomes

$$(9.14) \qquad p_{ij}(t) \geqq (1-\eta)\limsup_{\epsilon\to 0}\frac{p_{ij}(\epsilon)}{1-p_{ii}(\epsilon)},$$

for sufficiently small $t$. When $t\to 0$ this gives the first part of (9.2). Similarly if $i\neq j, j\in G_\alpha$

$$(9.13') \qquad p_{ji}(n\epsilon) \geqq \sum_{k=0}^{n-1} p_{jj}(\overline{n-k-1}\,\epsilon) p_{ji}(\epsilon) p_{ii}(\epsilon)^k \geqq (1-\eta)\frac{1-p_{ii}(\epsilon)^n}{1-p_{ii}(\epsilon)} p_{ji}(\epsilon)$$

is true for sufficiently small $n$, and then if $q_i=\infty$

$$(9.14') \qquad p_{ji}(t) \geqq (1-\eta)\limsup_{\epsilon\to 0}\frac{p_{ji}(\epsilon)}{1-p_{ii}(\epsilon)},$$

for sufficiently small $t$. When $t\to 0$ this gives the second part of (9.2). If $q_i<\infty$, (9.13) implies that

$$(9.15) \qquad p_{ij}(t) \geqq (1-\eta)\frac{1-e^{-q_i t}}{q_i}\limsup_{\epsilon\to 0}\frac{p_{ij}(\epsilon)}{\epsilon}.$$

Then

$$(9.16) \qquad \liminf_{t\to 0}\frac{p_{ij}(t)}{t} \geqq (1-\eta)\limsup_{\epsilon\to 0}\frac{p_{ij}(\epsilon)}{\epsilon}.$$

Since $\eta>0$ is arbitrary, this implies that $\lim_{t\to 0} p_{ij}(t)/t$ exists, and the limit is finite, from (9.15). Similarly equation (9.13′) implies that $\lim_{t\to 0} p_{ji}(t)/t$ exists and is finite. Moreover

$$(9.17) \qquad \sum_{j\neq i}\frac{p_{ij}(t)}{t} = \frac{1-p_{ii}(t)}{t},$$

so that (9.4) is true. Equation (9.17) can also be written

$$\sum_{j \neq i} \frac{p_{ij}(t)}{1 - p_{ii}(t)} = 1. \tag{9.18}$$

Then in the finite-dimensional case $q_i = \infty$ is impossible (since each term of the sum goes to 0 with $t$ if $q_i = \infty$), and (9.17) implies that there is equality in (9.4).

In discussing the continuity properties of $x(t)$ in $t$, it is usually convenient, because of measurability considerations, to choose a denumerable everywhere dense $t$-set $R$ and then consider the functions $x(r)$ for $r \in R$. The continuity properties of $x(r)$ can be interpreted as continuity properties of $x(t)$, if the proper space $\Omega$ of the stochastic process is chosen, and this will sometimes be done below.

THEOREM 10. *Suppose that hypothesis $H_\alpha$ is true. Let $\tau$ be any positive number and let $R$ be any denumerable everywhere dense set. Then* $\lim_{r \to \tau} x(r) = x(\tau)$ $(r \in R)$ *with probability* 1 *if and only if whenever $p_{\alpha i}(\tau) > 0$, $q_i$ is finite. If $q_i < \infty$ and if $p_{\alpha i}(\tau) > 0$, $q_{ij}/q_i$ is the conditional probability that if $x(\tau) = i$, and if there is a discontinuity of $x(r)$ $(r \in R)$ before $\tau + h$, then there is a first discontinuity before $\tau + h$, which is an isolated discontinuity where $x(r)$ jumps to $j$.*

The probability that $x(\tau) = i$ and that $x(r) = i$ for $r \in R$, $\tau - h < r < \tau + h$ is $p_{\alpha i}(\tau - h)\phi_i(2h)$. Then $\lim_{r \to \tau} x(r) = x(\tau)$ with probability 1 if and only if $\lim_{h \to 0} \phi_i(h) = 1$, that is, if and only if $q_i < \infty$, whenever $p_{\alpha i}(\tau) > 0$. This proves the first part of the theorem. The second part requires a more detailed analysis. Suppose that $q_i < \infty$ and that $p_{\alpha i}(\tau) > 0$. We shall evaluate the probability of the $x(t)$-set $\Lambda_\eta$ determined by the following conditions: $x(\tau) = i$; $x(r) = i$ for $r \in R$, $r > \tau$ on some interval of $r$-values; $x(r)$ then jumps to $j$, remaining equal to $j$ on some interval of length at least $\eta$, the jump occurring before $\tau + h$. Let $n$ be any positive integer, and define $\Lambda_{n,\eta}$ by

$$\begin{aligned}\Lambda_{n,\eta} = \sum_{m=1}^{n-2} \Big\{ x(\tau) = i;\ x(r) \equiv i, \tau < r < \tau + \frac{m}{n} h;\ x\left(\tau + \frac{m+1}{n} h\right) = j; \\ x(r) \equiv j, \tau + \frac{m+1}{n} h < r < \tau + \frac{m+1}{n} h + \eta \Big\}.\end{aligned} \tag{10.1}$$

Then

$$\begin{aligned} P(\Lambda_{n,\eta}) &= \sum_{m=1}^{n-2} p_{\alpha i}(\tau) e^{-mhq_i/n} p_{ij}(h/n) e^{-\eta q_i} \\ &= p_{\alpha i}(\tau) \frac{e^{-hq_i/n} - e^{-hq_i(1-1/n)}}{1 - e^{-hq_i/n}} p_{ij}(h/n) e^{-\eta q_i} \\ &\to p_{\alpha i}(\tau) \frac{1 - e^{-hq_i}}{q_i} q_{ij} e^{-\eta q_i}. \end{aligned} \tag{10.2}$$

Now if $x(t)\in\Lambda_\eta$ it follows that $x(t)\in\Lambda_{n,\eta'}$ for sufficiently large $n$, whenever $\eta'<\eta$:

$$\Lambda_\eta \subset \liminf_{n\to\infty} \Lambda_{n,\eta'}. \tag{10.3}$$

If $x(t)\in\Lambda_{n,\eta}$ for infinitely many values of $n$, $x(t)\in\Lambda_\eta$, if no $\tau+(m/n)h\in R$:

$$\limsup_{n\to\infty} \Lambda_{n,\eta} \subset \Lambda_\eta. \tag{10.4}$$

Then if $\eta'<\eta$

$$\limsup_{n\to\infty} P(\Lambda_{n,\eta}) \leqq P(\Lambda_\eta) \leqq \liminf_{n\to\infty} P(\Lambda_{n,\eta'}). \tag{10.5}$$

The inferior and superior limits in (10.5) are actual limits, evaluated in (10.2). Since the limit function of $\eta$ is continuous, we obtain[22], letting $\eta'\to\eta$,

$$P(\Lambda_\eta) = p_{\alpha i}(\tau)(1 - e^{-hq_i}) \frac{q_{ij}}{q_i} e^{-\eta q_j}. \tag{10.6}$$

The probability that $x(\tau)=i$, that $x(r)=i$ for $r>\tau$ on some $r$-interval ($r\in R$), and that then $x(r)$ jumps to $j$ where it remains for some $r$-interval, the jump occurring before $\tau+h$, is therefore

$$\lim_{\eta\to 0} P(\Lambda_\eta) = p_{\alpha i}(\tau)(1 - e^{-hq_i})q_{ij}/q_i, \tag{10.7}$$

and this equality is equivalent to the statement of the theorem.

To make clear the meaning of Theorem 10, suppose that $\sum_j q_{ij}=q_i<\infty$ for all $i$ in $G_\alpha$. Then if $\tau>0$, $\lim_{r\to\tau} x(r)=x(\tau)$ with probability 1. Excluding an $x(t)$-set of $\Omega^*$-measure 0, each $x(t)$ in the remainder $\Lambda$ is then equal on $R$ to $x(\tau)$ for $r$ sufficiently near $\tau$. According to the second part of the theorem, we can make the excluded $\Omega^*$-set so large that if $x(t)\in\Lambda$ there will be a first discontinuity of $x(r)$ (if any) after $\tau$, a jump. Now, applying the second part of the theorem, letting $\tau$ run through all rational numbers, we see that the excluded $\Omega^*$-set can be made so large that if $x(t)\in\Lambda$, there will be a second discontinuity (if there is more than one), also a jump, a third, and so on. These discontinuities may cluster at a point, to give $x(r)$ a discontinuity which is no longer a jump.

We shall use a somewhat indirect method in examining in more detail the transitions of the system, that is, the discontinuities of $x(t)$. This method has the advantage of exhibiting analytically the relation between the regularity of the matrix function $(p_{ij}(t))$ and the discontinuities of $x(t)$.

---

[22] We have tacitly assumed the measurability of $\Lambda_\eta$. This is easily proved directly, or the above discussion can be modified, using inner and outer measures in (10.5), to furnish the proof that $\Lambda_\eta$ is measurable, besides evaluating its measure. The restriction we have made on $h$ is essential for (10.4), but evidently (10.5) and (10.6) are true without this restriction.

Let $y_t$ (for each $t$ in some point set) be a chance variable. The family of chance variables $\{y_t\}$ will be said to have the property $\mathcal{E}$ if (for any natural number $n$), whenever $t_1 < \cdots < t_{n+1}$,

$$E\{y_{t_1}, \cdots, y_{t_n}; y_{t_{n+1}}\} = y_{t_n}(^{23}), \tag{11.1}$$

with probability 1(²⁴). Suppose a family $\{y_t\}$ has the property $\mathcal{E}$, for $t$ in some interval $(a, b)$. Then if $a<\tau<b$, $t_n \uparrow \tau$ implies that $\lim_{n\to\infty} y_{t_n} = y_{\tau-}$ exists with probability 1, and the limit $y_{\tau-}$ is independent (neglecting zero probabilities) of the particular sequence $\{t_n\}$. The chance variable $\{y_{\tau+}\}$ is defined similarly in terms of approach from above. Moreover, $y_{\tau-}=y_{\tau+}=y_\tau$ with probability 1, if $\tau$ is not in some set, which is at most denumerable. We shall call this set the set of fixed discontinuities. If $R$ is any denumerable set, dense on $(a, b)$, $y_r$ $(r\in R)$, with probability 1, considered as a function of $r$ alone is equal to a function defined on $(a, b)$, and continuous on the right at every point of $(a, b)$ not a fixed discontinuity point. It will be useful below to say that a family of chance variables $y_t$ has the property $\mathcal{E}^*$ if the family $y_{-t}$ has the property $\mathcal{E}$.

It is easily verified that if $T>0$, and if $y_t$ is defined by

$$y_t = p_{x(t)j}(T-t) \qquad (0<t<T), \tag{11.2}$$

then the family of chance variables has the property $\mathcal{E}$(²⁵). We shall show that there are no fixed discontinuities if hypothesis $H_\alpha$ is true. To do this it will be sufficient to show that if $t$ is given, and if $t_n \to t$, then some subsequence of $\{y_{t_n}\}$ converges to $y_t$ with probability 1. Since according to Theorem 8, $x_{t_n}(\omega)$ converges to $x_t(\omega)$ in measure, some subsequence, $x_{\tau_n}(\omega)$, converges to $x_t(\omega)$ with probability 1. Then

$$y_{\tau_n} = p_{x(t)j}(T-\tau_n) \tag{11.3}$$

for large $n$, with probability 1, so that $y_{\tau_n}\to y_t$ with probability 1, because the $p_{ij}(t)$ are continuous if $i\in G_\alpha$ and for each $t$, $P^*\{x_t(\omega)\in G_\alpha\}=1$. In a similar way it can be proved that if $T^*>0$, and if $y_t^*$ is defined by

$$y_t^* = \frac{p_{jx(t)}(t-T^*)}{p_{\alpha x(t)}(t)} \qquad (t>T^*), \tag{11.4}$$

the family of chance variables $\{y_t^*\}$ has the property $\mathcal{E}^*$, and there are no fixed discontinuities, if hypothesis $H_\alpha$ is true. The chance variable $y_t^*\cdot p_{\alpha j}(T^*)$

(²³) The notation $E\{y_{t_1}, \cdots, y_{t_n}; y_{t_{n+1}}\}$ will be used to denote the conditional expectation of $y_{t_{n+1}}$ for given values of $y_{t_1}, \cdots, y_{t_n}$, a function of the latter variables.

(²⁴) The properties of such a family, summarized here, are proved in the author's paper in these Transactions, vol. 47 (1940), pp. 455–486. This paper will be referred to as "$\mathcal{E}$."

(²⁵) This fact is a result of the well known relations between conditional expectation functions, and is a special case of the fact that if $\{w_t\}$ is any family of chance variables, if $z$ is a chance variable dependent on the $w_t$, and if $z_t=E\{w_s, s\leqq t; z\}$, ($z_t$=expectation of $z$ for $w_s$ given for $s\leqq t$), then the family $\{z_t\}$ has the property $\mathcal{E}$.

bears the same relation to the inverse process ($t$ decreasing) as $y_t$ bears to the given process. For each $t$, the denominator in (11.4) vanishes only with probability 0 (hypothesis $H_\alpha$). The following two regularity conditions on the $p_{ij}(t)$ will be useful.

Condition $C(\beta)$. *Let $\beta$ be in $G_\alpha$. Then there are numbers $\eta$, $\delta$ such that for all $i \neq \beta$ in $G_\alpha$, and all $s < \delta$*

$$p_{i\beta}(s) < 1 - \eta. \tag{11.5}$$

Condition $C^*(\beta, \tau)$. *Let $\beta$ be in $G_\alpha$ and let $\tau$ be a positive number with $p_{\alpha\beta}(\tau) > 0$. There are positive numbers $\eta$, $\delta$ such that if $0 < s < \delta$, $i \notin \beta$, $p_{\alpha i}(\tau) > 0$, then*

$$p_{\beta i}(s) < \frac{p_{\alpha i}(\tau + s)}{p_{\alpha\beta}(\tau)}(1 - \eta). \tag{11.5*}$$

Under hypothesis $H_\alpha$, if $i$ is fixed and $\delta \to 0$ in (11.5), the inequality becomes $0 \leqq 1 - \eta$, and under the same circumstances, (11.5*) becomes

$$0 \leqq \frac{p_{\alpha i}(\tau)}{p_{\alpha\beta}(\tau)}(1 - \eta). \tag{11.6}$$

Then conditions $C(\beta)$ and $C^*(\beta, \tau)$ are certainly always satisfied in the finite-dimensional case, under hypothesis $H_\alpha$, for all possible $\beta$ and pairs $\beta$, $\tau$ ($\beta \in G_\alpha$), respectively.

Condition $C(\beta)$ can be put in an interesting alternate form. If condition $C(\beta)$ is not satisfied, there is a sequence of distinct integers $\{i_\nu\}$ in $G_\alpha$, and a sequence $\{s_\nu\}$, $s_\nu \to 0$, such that $p_{i_\nu\beta}(s_\nu) \to 1$. Now if $t > 0$, and if $\nu$ is so large that $s_\nu < t$,

$$p_{i_\nu j}(t) = p_{i_\nu\beta}(s_\nu)p_{\beta j}(t - s_\nu) + \sum_{k \neq \beta} p_{i_\nu k}(s_\nu)p_{kj}(t - s_\nu). \tag{11.7}$$

If $j \notin G_\alpha$, $p_{i_\nu j}(t) \equiv 0$. If $j \in G_\alpha$, the sum on the right is at most

$$\sum_{k \neq \beta} p_{i_\nu k}(s_\nu) = 1 - p_{i_\nu\beta}(s_\nu) \to 0.$$

Then (11.7) implies

$$\lim_{\nu \to \infty} p_{i_\nu j}(t) = p_{\beta j}(t), \tag{11.8}$$

for all $j$, $t$. *Thus (under hypothesis $H_\alpha$) condition $C(\beta)$ is satisfied if (and, as is easily seen, only if) no sequence of distinct rows (whose elements have first subscripts in $G_\alpha$) converges, element by element to the $\beta$th row, for all $t$. An analogous but less elegant form of $C^*(\beta, \tau)$ can be obtained.*

The following theorem makes Theorem 10 more precise.

THEOREM 11. *Suppose that hypothesis $H_\alpha$ is true. Let $\tau$ be any positive number, and suppose that $R$ is any denumerable set having $\tau$ as a limit point. Then*

$$\lim_{r\to\tau} \frac{x(r) - x(\tau)}{1 + x(r)^2} = 0 \qquad (r \in R) \tag{11.9}$$

*with probability* 1. *If $p_{\alpha\beta}(\tau) > 0$, then $\lim_{r\to\tau} x(r) = \beta$ whenever $x(\tau) = \beta$, (with probability* 1) *if and only if $q_\beta < \infty$. If $C(\beta)$ or $C^*(\beta, \tau)$ is satisfied, then $q_\beta < \infty$.*

Since, as we have seen in Theorem 8, if $r \to \tau$, $x_r(\omega) \to x_\tau(\omega)$ in measure, that is, $x(r) \to x(\tau)$ in measure, it is impossible that $|x(r)| \to \infty$ with positive probability along any sequence of $r$-values approaching $\tau$. Therefore, neglecting 0 probabilities ($\Omega^*$-sets of measure 0), (11.9) implies that $x(r)$ always has $x(\tau)$ as a limiting value when $r \to \tau$, (the only finite limiting value) but $x(r)$ may also have $\pm\infty$ as a limiting value. In the course of the proof of Theorem 10, we have already proved that if $p_{\alpha\beta}(\tau) > 0$, $\lim_{r\to\tau} x(r) = \beta$ whenever $x(\tau) = \beta$ with probability 1, if and only if $q_\beta < \infty$: in fact this statement follows directly from our evaluation of $\phi_\beta(h)$. To prove (11.9) we shall use the families of chance variables $\{y_t\}$, $\{y_t^*\}$ introduced above. Since these families have no fixed discontinuities,

$$\lim_{r\to\tau} p_{x(r)j}(T - r) = p_{x(\tau)j}(T - \tau), \tag{11.10}$$

with probability 1. Let $\Lambda$ be an $\omega$-set of measure 1, such that the following conditions are satisfied, if $x(t) \in \Lambda$:

(a) (11.10) is true for all $j$ and rational $T > \tau$;

(b) if $p_{\alpha j}(r) = 0$, then $x(r) \neq j$; if $p_{\alpha j}(\tau) = 0$, then $x(\tau) \neq j$.

Now suppose that $x_0(t) \in \Lambda$, and suppose that $x_0(\tau) = \beta$. Suppose that $\liminf_{r\to\tau} |x_0(r)| < +\infty$. Then there is an integer $\gamma$ such that $x_0(r) = \gamma$ for infinitely many values of $r$, as $r \to \tau$. Condition (a) implies

$$\lim_{r\to\tau} p_{\gamma j}(T - r) = p_{\gamma j}(T - \tau) = p_{\beta j}(T - \tau), \tag{11.11}$$

for all $j$ and rational $T > \tau$. (We are using the fact that because of condition (b), $p_{\alpha\gamma}(t) \not\equiv 0$, so, in accordance with hypothesis $H_\alpha$, $p_{\alpha\gamma}(t)$ is continuous in $t$.) Because of hypothesis $H_\alpha$, $\lim_{t\to 0} p_{\gamma j}(t) = \delta_{\gamma j}$, $\lim_{t\to 0} p_{\alpha j}(t) = \delta_{\alpha j}$ for all $j$. Thus if we let $T \to \tau$, (11.11) implies that $\delta_{\gamma j} = \delta_{\beta j}$ for all $j$: $\gamma = \beta$. We have now proved that the only possible limiting values of $x_0(r)$ as $r \to \tau$ are $x_0(\tau)$, $\pm\infty$; (11.9) is true for $x(t) \in \Lambda$, and hence is true with probability 1. If the matrices are finite-dimensional, there can be only finite limiting values of $x_0(r)$; so $x(r) \to x(\tau)$ $(r \in R)$ with probability 1, as we have already proved in Theorem 10. Now suppose that $(p_{ij}(t))$ is infinite-dimensional, and suppose that for some $x_0(t)$ in $\Lambda$, $x_0(r)$ does not approach $x_0(\tau) = \beta$ $(r \in R)$. Then there must be a sequence of integers $\{i_\nu\}$ and a sequence $\{r_\nu\}$ such that $x_0(r_\nu) = i_\nu \to \pm\infty$, $(r_\nu \to \tau)$. Then (11.10) becomes

$$\lim_{\nu\to\infty} p_{i_\nu j}(T - r_\nu) = p_{\beta j}(T - \tau) \tag{11.12}$$

for all $j$ and rational $T > \tau$. It follows readily from (11.12) with $j = \beta$ that there is a subsequence $\{j_\nu\} = \{i_{a_\nu}\}$ of $\{i_\nu\}$ and a sequence of values $\{T_\nu\}$ of $T$, $T_\nu \downarrow \tau$, such that

$$\lim_{\nu\to\infty} p_{j_\nu\beta}(T_\nu - r_{a_\nu}) = \lim_{\nu\to\infty} p_{\beta\beta}(T_\nu - \tau) = 1. \tag{11.13}$$

This evidently contradicts condition $C(\beta)$. Thus if $C(\beta)$ is satisfied, $q_\beta < \infty$. The family of chance variables

$$\left\{\frac{p_{jx(r)}(r - \tau)}{p_{\alpha x(r)}(r)}\right\} \qquad (r > \tau)$$

has, as we have seen, the property $\mathcal{E}^*$. Since these chance variables are non-negative, there is convergence when $r \downarrow \tau$, with probability 1([28]). Since $x(r) \to x(\tau)$ in measure,

$$\lim_{r\downarrow\tau} \frac{p_{\beta x(r)}(r - \tau)}{p_{\alpha x(r)}(r)} = \frac{1}{p_{\alpha\beta}(\tau)} \tag{11.14}$$

almost everywhere where $x(\tau) = \beta$. We can suppose $\Lambda$ has been chosen so that (11.14) is true if $x(t) \in \Lambda$. Unless $x(r) \to \beta$ whenever $x(t) \in \Lambda$, $r \downarrow \tau$, and $x(\tau) = \beta$, there is an $x_0(t)$ in $\Lambda$ with $x_0(\tau) = \beta$, a sequence of integers $\{i_\nu\}$, and a sequence $\{r_\nu\}$ such that $x_0(r_\nu) = i_\nu \to \pm\infty$, $r_\nu \downarrow \tau$. Then, using (11.14),

$$\frac{p_{\beta i_\nu}(r_\nu - \tau)}{p_{\alpha i_\nu}(r_\nu)} \to \frac{1}{p_{\alpha\beta}(\tau)} \tag{11.15}$$

which contradicts $C^*(\beta, \tau)$. Thus if $C^*(\beta, \tau)$ is satisfied, $q_\beta < \infty$. (We are using here the fact which is implicit in the discussion of $\phi_\beta(h)$ above that if $x(\tau) = \beta$, $\lim_{r\downarrow\tau} x(r) = \beta$ with probability 1 if and only if $q_\beta < \infty$.)

We shall need a somewhat stronger condition than $C^*(\beta, \tau)$ below. We shall say that condition $C^{**}(\beta, \tau)$ is satisfied if $p_{\alpha\beta}(\tau) > 0$ and if there are positive numbers $\eta$, $\delta$ such that if $0 < s_1 \leq s_2 < \delta$, $i \neq \beta$, $p_{\alpha i}(\tau + s_1) > 0$, then

$$p_{\beta i}(s_2) < \frac{p_{\alpha i}(\tau + s_1)}{p_{\alpha\beta}(\tau)}(1 - \eta). \tag{11.16}$$

Condition $C^{**}(\beta, \tau)$ is always satisfied in the finite-dimensional case, under hypothesis $H_\alpha$, if $\beta \in G_\alpha$ since (11.16) becomes (11.6) when $s_1$ and $s_2$ approach 0.

---

([28]) This is true if $r \downarrow \tau$ along any sequence of values (Doob, these Transactions, vol. 47 (1940), p. 460, Theorem 1.3) and this means the truth of the statement when $r \downarrow \tau$, $r \in R$ (Doob, these Transactions, vol. 42 (1937), p. 111, Theorem 1.3, or, in another formulation, Duke Mathematical Journal, vol. 4 (1938), pp. 758–759, Lemma 2).

THEOREM 12. *Suppose that hypothesis $H_\alpha$ is true. Let $R$ be any denumerable everywhere dense $t$-set. Then there is a set $\Lambda$ of functions $x(t)$, of probability* 1, *such that if $x(t)\in\Lambda$ the following statements are true.*

(a) *If the matrices are finite-dimensional, $x(r)$ $(r\in R)$ has only isolated jumps as discontinuities.*

(b) *Either* $\lim_{r\downarrow\tau}|x(r)|=\infty$ $(r\in R)$, *or there is an integer $\beta$, depending on the function $x(t)$ and on $\tau$, such that*

$$\lim_{r\downarrow\tau}\frac{x(r)-\beta}{1+|x(r)|^2}=0 \qquad (r\in R). \tag{12.1}$$

*If condition $C(\beta)$ is satisfied, then either* $\lim_{r\downarrow\tau}|x(r)|=\infty$ *or there is an integer $\beta$ depending on the function $x(t)$ and on $\tau$, such that* $\lim_{r\downarrow\tau}x(r)=\beta$ $(r\in R)$.

(b*) *The statement of* (b) *remains true with $r\uparrow\tau$, instead of $r\downarrow\tau$, replacing condition $C(\beta)$ by $C^{**}(\beta,\tau)$.*

(c) *For each $\tau$, there will be an integer $\beta$ (and $\beta=x(\tau)$), as described in* (b), (b*), *with probability* 1.

Theorem 12 is closely related to work of Doeblin and Feller, with which it will be compared below.

The families $\{y_t\}$, $\{y_t^*\}$ have the properties $\mathcal{E}$, $\mathcal{E}^*$, respectively. There is therefore an $\omega$-set $\Lambda$ of probability 1, such that if $x(t)\in\Lambda$, the corresponding $y_t$, $(y_t^*)$ coincide on $R$ with functions everywhere continuous on the right (left), for all $j$, rational $T$, $T^*$. (It has been proved above that there are no points of fixed discontinuity.) We can also suppose that $x(r)\neq j$ unless $p_{\alpha j}(r)>0$, if $x(t)\in\Lambda$. Then if $x_0(t)\in\Lambda$, and if $T>\tau>T^*>0$, $T$, $T^*$ rational, the following limits exist:

$$\lim_{r\downarrow\tau} p_{x_0(r)j}(T-r), \tag{12.2}$$

$$\lim_{r\uparrow\tau}\frac{p_{jx_0(r)}(r-T^*)}{p_{\alpha x_0(r)}(r)}\cdot \tag{12.2*}$$

If $x_0(r)$ takes on a subscript $\beta$ for values of $r$ approaching $\tau$ from above, we can evaluate the limit in (12.2):

$$\lim_{r\downarrow\tau} p_{x_0(r)j}(T-r)=\lim_{r\downarrow\tau}p_{\beta j}(T-r)=p_{\beta j}(T-\tau). \tag{12.3}$$

Then $\beta$ is uniquely determined, for if $\gamma$ had the same defining property, we should have $p_{\beta j}(T-\tau)=p_{\gamma j}(T-\tau)$ for all rational $T>\tau$, where $\beta$, $\gamma$ are both in $G_\alpha$. When $T\downarrow\tau$ this means (using hypothesis $H_\alpha$) that $\delta_{\beta j}=\delta_{\gamma j}$ for all $j$, impossible unless $\beta=\gamma$. Thus (12.1) is true. Then in the finite-dimensional case, $\lim_{r\downarrow\tau}x_0(r)=\beta$, that is, $x_0(r)=\beta$ for $r$ sufficiently near $\tau$ $(r>\tau)$. In the infinite-dimensional case, unless $\lim_{r\downarrow\tau}x_0(r)=\beta$, we have proved that $\limsup_{r\downarrow\tau}|x_0(r)|=\infty$, and the method of proof of Theorem 11 can be car-

ried through to find a contradiction to condition $C(\beta)$. We have now proved that Theorem 12(b) is true, supposing however that there is a subscript $\beta$ as described. (In the finite-dimensional case there is always such a $\beta$.) In the infinite-dimensional case, if there is no such $\beta$, $\lim_{r\downarrow\tau} |x(r)| = \infty$. This finishes the proof of (b). The discussion when $r \uparrow \tau$ is carried on in the same way, using the existence of the limit in (12.2*). Theorem 12(a) is now obviously true. For a given $\tau$, $x(r)\rightarrow x(\tau)$ in measure, when $r\rightarrow\tau$, so there will be an integer $\beta$ as described above, and $\beta = x(\tau)$, with probability 1.

As usual in this sort of discussion, instead of saying that $x(r)$ $(r\in R)$ has the above described properties with probabiliy 1, we could say that if a space $\Omega$ of a stochastic process is chosen properly, all the $x(t)$ in $\Omega$ will have the above properties, where $t$ ranges through all values.

Doeblin has considered a general Markoff process in which the transition probability of going from state $i$ at time $t$ to state $j$ at time $t'$ is not supposed necessarily to be a function of $t'-t$, and in which it is not supposed that the number of possible states is denumerably infinite. His hypotheses, when translated into our notation, and simplified because of the more special process being considered here, become

$$\lim_{t\to 0} p_{ii}(t) = 1 \tag{12.4}$$

uniformly in $i$. This hypothesis, combined with the hypothesis that the process is initially in state $\alpha$ is considerably stronger than hypothesis $H_\alpha$ (except in the finite-dimensional case, when, assuming hypothesis $H_\alpha$, Doeblin's condition is always applicable) and evidently also implies condition $C(\beta)$ for all $\beta\in G_\alpha$. Doeblin showed that under his hypotheses, and assuming some given initial state, neglecting an $\omega$-set of measure 0, $x(r)$ $(r\in R)$ has only isolated jumps as discontinuities([27]).

Conversely, suppose that the process is initially in state $\alpha$, and that the $\omega$-measure has the property that $x(r)$ ($r\in R$, a denumerable everywhere dense $t$-set) has only isolated jumps as discontinuities, with probability 1. Theorem 11 shows that in this case there must be continuity at each fixed $\tau$, with probability 1, that is, $q_i<\infty$ if $i\in G_\alpha$. Also, by Theorem 10, $\sum_j q_{ij}=q_i$. Let $P_{ij}^{(n)}(t)$ be the probability([28]) that if $x(\tau)=i$ then $x(\tau+t)=j$ and $x(r)$ has $n$ jumps in going from $i$ to $j$, between $\tau$ and $\tau+t$. It is easily verified that if $i\in G_\alpha$

$$\begin{aligned} P_{ik}^{(0)}(t) &= \delta_{ik}e^{-q_i t} \\ P_{ik}^{(n+1)}(t) &= \sum_j \int_0^t P_{ij}^{(n)}(s)q_{jk}e^{-q_k(t-s)}ds \qquad (n\geqq 0), \end{aligned} \tag{12.5}$$

([27]) Skandinavisk Aktuarietidskrift, vol. 22 (1939), pp. 211–222.

([28]) For this conditional probability to have a meaning we must suppose that $p_{\alpha i}(\tau)>0$; $\tau$ can always be so chosen, if $i\in G_\alpha$.

and obviously

$$p_{ij}(t) = \sum_{n\geq 0} P_{ij}^{(n)}(t). \tag{12.6}$$

Moreover if we suppose only hypothesis $H_\alpha$ to be true, and that $\sum_j q_{ij} = q_i < \infty$, $(i \in G_\alpha)$, then if $i \in G_\alpha$, considerations analogous to those used in the proof of Theorem 10 show that $P_{ij}^{(n)}(t)$ as defined in (12.5) will have the probability meaning described above. On the other hand, (12.6) is now true if and only if the only discontinuities of $x(r)$ are isolated jumps, with probability 1. Feller has found necessary and sufficient conditions on the $q_i$, $q_{ij}$ that (12.6) be true([29]). The above remarks give a complete justification for Feller's probability interpretation (ibid., p. 498) of the $P_{ij}^{(n)}(t)$. (He did not need this interpretation in his proofs.) The details given in Theorem 12 on the character of $x(t)$ at a discontinuity which is not a jump round out Feller's description, arising from an entirely different background (ibid., pp. 512–513).

Suppose again that hypothesis $H_\alpha$ is satisfied and that $\sum_j q_{ij} = q_i < \infty$ for all $i \in G_\alpha$. The differential equations

$$p'_{ik}(t) = -q_i p_{ik}(t) + \sum_{j\neq i} q_{ij} p_{jk}(t) \qquad (i \in G_\alpha) \tag{13.1}$$

satisfied by the $p_{ij}(t)$, due to Kolmogoroff([30]) are well known, but their intimate relation to the continuity properties of the $x(t)$ seems less well known. In fact, under the above hypotheses, we have shown (Theorem 10) that (with probability 1) if $x(\tau) = i \in G$, there is a first discontinuity of $x(r)$ ($r \in R$, an everywhere dense denumerable $t$-set) after $\tau$, an isolated jump. Then the probability of going from $i$ to $k$ is the sum of the probabilities of going from $i$ to $j$ on the first jump, and then to $k$ (summed over $j$). Considerations analogous to those used in the proof of Theorem 10 now show that, evaluating the above probabilities,

$$p_{ik}(t) = \sum_{j\neq i} \int_0^t e^{-q_i(t-s)} q_{ij} p_{jk}(s) ds + \delta_{ik} e^{-q_i t}. \tag{13.2}$$

The equations of (13.1) are obtained by differentiating those of (13.2), in which the series can obviously be differentiated term by term. The second set of differential equations obtained by Kolmogoroff

$$p'_{ik}(t) = -q_k p_{ik}(t) + \sum_{j\neq k} p_{ij}(t) q_{jk} \qquad (i \in G_\alpha) \tag{13.3}$$

---

([29]) These Transactions, vol. 48 (1940), pp. 506–507. Feller's results are applicable to considerably more general stochastic processes than those considered here.

([30]) Mathematische Annalen, vol. 104 (1931), p. 429. Kolmogoroff imposed further restrictions on the $p_{ij}(t)$. Feller (op. cit., p. 495) obtained (13.1) with substantially our hypotheses given above.

does not seem always to be true without further hypotheses. We shall show that the truth of (13.3) is equivalent to the imposition of certain regularity properties on the $x(t)$. The probability

$$p_{ik}(t_2) - p_{ik}(t_1)e^{-q_k(t_2-t_1)} \qquad (i \neq k,\ t_2 > t_1)$$

is at least equal to the probability that if $x(\tau-t_2)=i$, then $x(r)$ goes to $j$ at some point between $\tau-t_1$ and $\tau$, when $x(r)$ jumps to $k$, remaining at $k$ until $t=\tau$, summed over $j\neq k$. Thus

$$(13.4) \qquad p_{ik}(t_2) - p_{ik}(t_1)e^{-q_k(t_2-t_1)} \geqq \sum_{j\neq k}\int_{t_1}^{t_2} p_{ij}(s)q_{jk}e^{q_k(t-t_2-s)}ds.$$

Moreover there is equality if and only if when $x(\tau)=i$, there is a last discontinuity of $x(r)$ before $\tau$, which is a jump, with probability 1. Dividing (13.4) by $t_2-t_1$ and letting $t_2-t_1\to 0$ we obtain

$$(13.5) \qquad p'_{ik}(t) \geqq -q_k p_{ik}(t) + \sum_j p_{ij}(t)q_{jk} \qquad (i \in G_\alpha).$$

It is easily verified that (13.5) also follows directly from (0.1). Since there is equality in (13.5) if and only if there is equality in (13.4), we have obtained the following theorem.

THEOREM 13. *Suppose that hypothesis $H_\alpha$ is satisfied and that $\sum_j q_{ij}=q_i<\infty$ for all $i\in G_\alpha$. Then* (13.1) *is always true*; (13.3) *is true (for all $t$) if and only if when $x(\tau)=i$, there is, with probability* 1, *a last discontinuity of $x(r)$ before $\tau$, which is an isolated discontinuity (a jump).*

It is interesting to note that if $\bar{p}_{ij}(t)$ is the probability that if $x(\tau)=i$ then $x(\tau+t)=j$ and the transition from $i$ to $j$[31] is accomplished in a finite number of isolated jumps, then $\bar{p}_{ij}(t)$ evidently satisfies (0.1) except that $\sum_j \bar{p}_{ij}(t)$ may be less than 1. Moreover (13.1) is also true for the $\bar{p}_{ij}(t)$ since the derivation for the $p_{ij}(t)$ applies equally well to $\bar{p}_{ij}(t)$. And the derivation we have given of (13.5), when applied to the $\bar{p}_{ij}(t)$ actually gives equality: (13.3) is true of the $\bar{p}_{ij}(t)$ in all cases. The latter fact was also proved by Feller.

(31) Strictly speaking, we should restrict $i, j$ to lie in $G_\alpha$.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

# ON CONVERSE GAP THEOREMS

BY

GEORGE PÓLYA

1. In what follows, I consider power series with preassigned vanishing coefficients. I write such a series in the form

$$(1) \qquad a_1z^{\lambda_1} + a_2z^{\lambda_2} + \cdots + a_nz^{\lambda_n} + \cdots.$$

The numbers $a_n$ are different from 0 and the $\lambda_n$ are integers,

$$(2) \qquad 0 < \lambda_1 < \lambda_2 < \cdots < \lambda_n < \lambda_{n+1} < \cdots.$$

I assume that the radius of convergence of the series (1) is finite and different from 0.

I quote two well known theorems[1].

THEOREM I. *If*

$$(3) \qquad \lim_{n\to\infty} n\lambda_n^{-1} = 0$$

*the domain of existence of the analytic function defined by* (1) *is the interior of the circle of convergence of the series* (1).

THEOREM II. *If*

$$(4) \qquad \liminf_{n\to\infty} n\lambda_n^{-1} = 0$$

*the domain of existence of the analytic function defined by* (1) *is a simply connected part of the z-plane* (*from which it follows that the function defined by* (1) *is uniform*).

Is it possible to improve these theorems by enlarging the hypothesis? Is there a less exacting hypothesis leading to the same conclusion? I say no. In fact I shall solve the following problems[2].

---

Presented to the Society, February 22, 1941; received by the editors June 4, 1941.

(1) In the following the two parts of my paper *Untersuchungen über Lücken und Singularitäten von Potenzreihen*, Mathematische Zeitschrift, vol. 29 (1929), pp. 549–640 and Annals of Mathematics, (2), vol. 34 (1933), pp. 731–777 will be quoted as LS I and LS II. For Theorem I (Fabry's theorem) see LS I, p. 627, Theorem VIa; for Theorem II (theorem of the present author) see LS II, p. 737, Theorem B.

(2) These problems have been stated, with an indication of the proof: I in Comptes Rendus de l'Académie des Sciences, Paris, vol. 208 (1939), pp. 709–711; II in the Bulletin of the American Mathematical Society, vol. 47 (1941), p. 207. A problem related to I was stated by G. Szegö, Acta Litterarum ac Scientiarum Szeged, vol. 1 (1923), p. 73.

PROBLEM I. *Given an infinite sequence of integers* $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ *satisfying* (2) *but not satisfying* (3), *find a power series of the form* (1) *defining an analytic function whose domain of existence extends beyond the circle of convergence.*

PROBLEM II. *Given an infinite sequence of integers* $\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots$ *satisfying* (2) *but not satisfying* (4), *find a power series of the form* (1) *defining a multiform analytic function* (*whose domain of existence, consequently, cannot be a simply connected part of the z-plane*).

2. It is of some interest to restate the facts which we proposed to prove in a different terminology. This terminology is due to Borel but, so far as I know, it was used by him just once([3]) and it has never been used since.

Let us consider power series whose radius of convergence is different from 0 and $\infty$ and let us say that two such series

$$\sum_0^\infty c_n z^n, \qquad \sum_0^\infty c_n' z^n$$

belong to the same class, if they have the same distribution of nonvanishing coefficients, that is if, for $n=0, 1, 2, \cdots$, $c_n$ and $c_n'$ are either both 0 or both different from 0. In fact, a class of power series is characterized by those powers of $z$ whose coefficients do not vanish, and therefore by an increasing sequence of integers $\lambda_1, \lambda_2, \cdots$ and all series belonging to the class have the same form (1)([4]). Let us call

$$\lim_{n\to\infty} n\lambda_n^{-1}, \qquad \liminf_{n\to\infty} n\lambda_n^{-1}, \qquad \limsup_{n\to\infty} n\lambda_n^{-1}$$

(in the given order) the *density*, the *lower density* and the *upper density* of the class (the first may not exist, but the second and the third necessarily exist). With this terminology, we may compound Theorem I and Problem I into one short statement (and the same for Theorem II and Problem II):

I. *In order that all power series of a class have their circle of convergence as natural boundary it is necessary and sufficient that the density of the class be* 0.

II. *In order that all power series of a class define uniform analytic functions it is necessary and sufficient that the lower density of the class be* 0.

A few other well known facts on the singularities of power series may be quite elegantly stated in the same terminology([5]).

---

([3]) Comptes Rendus de l'Académie des Sciences, Paris, vol. 137 (1903), pp. 695–697.

([4]) Series (1) corresponds to the (evidently unessential) assumption that the coefficient of $z^0$ is zero in all series of the class.

([5]) See also, for further literature, LS I, p. 622, Theorem IIIa for the first, and LS II, p. 745, Theorem IV for the third statement. The second statement is easy.

*Each class contains non-continuable power series.*

*In order that a class contain a power series having on its circle of convergence a pole and no other singular point, it is necessary and sufficient, that the sequence* $\lambda_1, \lambda_2, \cdots$ *contain nearly all integers, that is, either all integers or all with a finite number of exceptions.*

*In order that a class contain a power series having on its circle of convergence an essential singular point* (an isolated singularity which is not a pole and not a branch point) *and no other singular point, it is necessary and sufficient that the density of the class exist and be* 1.

These theorems seem to suggest that there are a few more of the same kind.

3. Our solution of the Problems I and II uses some properties of the series

$$F(1)z + F(2)z^2 + \cdots + F(m)z^m + \cdots = \Phi(z) \tag{5}$$

where $F(z)$ denotes an entire function of exponential type[6]. The real-valued periodic function $h(\phi)$ of the real variable $\phi$, defined by

$$h(\phi) = \limsup_{r\to\infty} r^{-1} \log |F(re^{i\phi})| \tag{6}$$

is called the indicator of $F(z)$. The series (5) has a finite radius of convergence and its analytic continuation is closely connected with the indicator $h(\phi)$. I quote the following fact[7]:

*If*

$$h(-\pi/2) < \pi, \qquad h(\pi/2) < \pi \tag{7}$$

*then the function* $\Phi(z)$ *defined by the series* (5) *is regular along the negative real axis, and also at the point* $z = \infty$ *in whose neighborhood it has the development*

$$\Phi(z) = -F(0) - \frac{F(-1)}{z} - \frac{F(-2)}{z^2} - \cdots. \tag{8}$$

This theorem yields a quick solution of Problem II. In fact, assume that the sequence of integers $\lambda_1, \lambda_2, \cdots$ does not satisfy (4). Then there exists a positive $\alpha$, such that

$$\frac{n}{\lambda_n} > \alpha. \tag{9}$$

Define

$$G(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z^2}{\lambda_n^2}\right). \tag{10}$$

---

(6) Defined in LS I, p. 578.

(7) See LS I, pp. 604–609, especially formulae (72) and (73), p. 609. Observe that the $\Phi(z)$ of these formulae differs by an additive constant from the $\Phi(z)$ of the present formula (5).

It follows from (9), (10) that, for $r>0$,

$$(11)\qquad G(ir)=\prod_{n=1}^{\infty}\left(1+\frac{r^2}{\lambda_n^2}\right)>\prod_{n=1}^{\infty}\left(1+\frac{\alpha^2r^2}{n^2}\right)=\frac{\sin i\pi\alpha r}{i\pi\alpha r}=\frac{e^{\pi\alpha r}-e^{-\pi\alpha r}}{2\pi\alpha r}.$$

Define $F(z)$ by the equation

$$(12)\qquad \pi z F(z)G(z)=\sin \pi z.$$

$F(z)$ is evidently an entire function of exponential type[8]. For positive integral $m$

$$(13)\qquad F(m)=\begin{cases}(-1)^{\lambda_n}/(\lambda_n G'(\lambda_n)) & \text{if}\quad m=\lambda_n\\ 0 & \text{if}\quad m\neq\lambda_n\end{cases}$$

$$(14)\qquad F(0)=1,$$

and, by virtue of (11) and (12),

$$(15)\qquad \limsup_{r\to\infty} r^{-1}\log|F(\pm ir)|\leqq \pi-\pi\alpha.$$

Consider the power series, arising from (5),

$$(16)\qquad \int_0^z \Phi(z)z^{-1}dz=\sum_1^{\infty}\frac{F(m)z^m}{m}=\sum_{n=1}^{\infty}\frac{(-1)^{\lambda_n}z^{\lambda_n}}{\lambda_n^2G'(\lambda_n)}.$$

Series (16) is evidently of form (1). The condition (7) is fulfilled, see (6) and (15). Therefore, the analytic continuation of (16) along the negative real axis is possible. But, in a certain neighborhood of the point $z=\infty$, we have, by virtue of (8) and (14), with a certain constant $C$,

$$\int_0^z \Phi(z)z^{-1}dz=C-\log z+\frac{F(-1)}{z}+\frac{F(-2)}{2z^2}+\cdots$$

and therefore the analytic continuation of (16) is not a single-valued function. Thus (16) fulfills all the requirements of Problem II.

4. Now we are going to solve Problem I. We use again the series (5) and the connection between the analytic continuation of this series and the indicator $h(\phi)$. We need now the following facts[9].

*To each entire function $F(z)$ of exponential type corresponds a bounded and closed convex domain $\mathfrak{Z}$, called the indicator diagram of $F(z)$. The domain $\mathfrak{Z}$ lies in the half-plane*

$$x\cos\phi+y\sin\phi-h(\phi)\leqq 0$$

(8) In fact, (24) holds also for the present $F(z)$.

(9) See LS I, pp. 604–609, especially p. 606. (Correct the misprint at the end of the fourteenth line from the top of p. 606; read $h(0)$ instead of 0.) I write as usual $z=x+iy$ with real $x$, $y$ in the following statement and later in (26), and (27).

*but has a point on the boundary of this half-plane; this holds for all real $\phi$. If the circle of convergence of the power series* (5) *is a natural boundary, the boundary of the indicator diagram $\mathfrak{J}$ of $F(z)$ contains a segment of a vertical straight line, of length not less than $2\pi$, limiting $\mathfrak{J}$ from the right.*

We shall use this fact at the end of the following construction which we divide in successive steps.

i. We start from a sequence $\lambda_1, \lambda_2, \cdots$ that does not satisfy (3). It follows that, for a fixed $\alpha$, $0<\alpha<1$, there are arbitrarily distant intervals of the form $r\alpha<x<r$ which contain more than $r\delta$ points of the sequence $\lambda_1, \lambda_2, \cdots, \delta$ being positive, sufficiently small, but fixed([10]). Thus we can choose a sequence of increasing positive integers $n_1, n_2, \cdots$ and a positive $\delta$, satisfying the following condition:

(I) The interval between $(n_k-1/2)/2^{1/2}$ and $n_k-1/2$, which I denote by $I_k$, contains at least $\delta n_k$ points of the sequence $\lambda_1, \lambda_2, \cdots$.

By rejecting if necessary certain elements of the chosen sequence, we can find a sequence (a subsequence of the first chosen sequence, which, by an appropriate change of notation, will be called again $n_1, n_2, \cdots$) satisfying not only condition (I) but also the following:

$$n_{k-1} < (n_k - \tfrac{1}{2})/2^{1/2},$$

$$\text{(II)} \qquad \log\left[\left(\frac{n_k}{1}\right)^2 - 1\right]\left[\left(\frac{n_k}{2}\right)^2 - 1\right] \cdots \left[\left(\frac{n_k}{n_{k-1}}\right)^2 - 1\right] < (n_k - \tfrac{1}{2})^{1/2}.$$

In fact, $n_{k-1}$ being fixed, both inequalities (II) are satisfied by any sufficiently great $n_k$.

Call those elements of the sequence $\lambda_1, \lambda_2, \cdots$ which are contained in the intervals $I_1, I_2, \cdots, I_k, \cdots$, numbered in order of magnitude, $\mu_1, \mu_2, \mu_3, \cdots$.

ii. Using the sequence $\mu_1, \mu_2, \cdots$ we just constructed, define

$$G(z) = \prod_{n=1}^{\infty}\left(1 - \frac{z^2}{\mu_n^2}\right). \tag{17}$$

This definition is different from (10), which we used in solving Problem II, and which we disregard now. We shall estimate

$$\begin{aligned}(n_k - \tfrac{1}{2})^{-1}\log\left|G(n_k - \tfrac{1}{2})\right| &= \sum_{j=1}^{\infty}(n_k - \tfrac{1}{2})^{-1}\log\left|\left(\frac{n_k - \frac{1}{2}}{\mu_j}\right)^2 - 1\right| \\ &= S_1 + S_2 + S_3.\end{aligned} \tag{18}$$

$S_2$ contains those terms of the sum on the right-hand side of the first line whose $\mu_j$ is contained in $I_k$; $S_1$ those terms whose $\mu_j$ is contained in one of

([10]) The easy proof can be left to the reader; it is contained in the fuller developments of LS I, pp. 556–560, especially in (14) p. 559.

the intervals $I_1, I_2, \cdots, I_{k-1}$; $S_3$ those whose $\mu_j$ is in one of the intervals $I_{k+1}, I_{k+2}, \cdots$. These intervals do not overlap, by the first inequality (II). If $\mu_j$ is in $S_1$ then $\mu_j < n_{k-1} - 1/2$ and therefore, by the second inequality (II),

$$S_1 < (n_k - \tfrac{1}{2})^{-1/2} \to 0 \tag{19}$$

as $n_k \to \infty$. If $\mu_j$ is in $S_3$ then $\mu_j > n_k - 1/2$ and therefore each term in $S_3$ is negative,

$$S_3 < 0. \tag{20}$$

The terms in $S_2$ are also negative, and in number not less than $n_k\delta$, by condition (I); the single term decreases algebraically (increases in absolute value) as $\mu_j$ increases. We obtain the (algebraically) greatest value of $S_2$ by taking as few terms as possible and terms as close to the left-hand end point of $I_k$ as possible. If the summation is extended to the integers $l$ satisfying.

$$(n_k - \tfrac{1}{2})/2^{1/2} < l < (n_k - \tfrac{1}{2})/2^{1/2} + n_k\delta$$

we have

$$\begin{aligned} S_2 &< \sum (n_k - \tfrac{1}{2})^{-1} \log \left[\left(\frac{n_k - \frac{1}{2}}{l}\right)^2 - 1\right] \\ &\to \int_{1/2^{1/2}}^{(1/2^{1/2})+\delta} \log\left(\frac{1}{t^2} - 1\right) dt < 0. \end{aligned} \tag{21}$$

By (18), (19), (20), (21) we obtain, $n$ running through the positive integers, that

$$\liminf_{n\to\infty} (n - \tfrac{1}{2})^{-1} \log \left| G(n - \tfrac{1}{2}) \right| < 0. \tag{22}$$

iii. Define $F(z)$ by (12). From (12) and (22) it follows that

$$\begin{aligned} h(0) &= \limsup_{r\to\infty} r^{-1} \log \left| F(r) \right| \\ &\geqq \limsup_{n\to\infty} (n - \tfrac{1}{2}) \log \frac{1}{\pi(n - \frac{1}{2}) \left| G(n - \frac{1}{2}) \right|} > 0. \end{aligned} \tag{23}$$

On the other hand, if $|z| = r$,

$$\left| F(z) \right| \leqq F(ir) < \prod_{n=1}^{\infty} \left(1 + \frac{r^2}{n^2}\right) = \frac{e^{\pi r} - e^{-\pi r}}{2\pi r} \tag{24}$$

and therefore, by (6), for all real $\phi$

$$h(\phi) \leqq \pi. \tag{25}$$

Inequalities for $h(\phi)$ are equivalent to geometric conditions for $\mathfrak{J}$, the indicator diagram of $F(z)$. In the present case, $F(z)$ is an even function and real-

valued for real $z$; so $\mathfrak{J}$ is symmetrical with respect to the real and to the imaginary axes. By (25), $\mathfrak{J}$ is contained in the circle

$$x^2 + y^2 = \pi^2. \tag{26}$$

By (23), it has a common point with the vertical line

$$x = h(0) \tag{27}$$

and no point with an abscissa greater then $h(0)$. But the segment, intercepted on the line (27) by the circle (26) is shorter than $2\pi$, because, by (23), $h(0) > 0$; *there is no segment of a vertical line, of length greater than or equal to* $2\pi$, *on the boundary of the indicator diagram* $\mathfrak{J}$.

iv. This last result shows that the series (5)

$$\Phi(z) = \sum_1^\infty F(m)z^m = \sum_{n=1}^\infty \frac{(-1)^{\mu_n} z^{\mu_n}}{\mu_n G'(\mu_n)} \tag{28}$$

is *continuable*, by virtue of the fact quoted at the beginning of this section. Series (28) is not exactly of the form (1), because some of the coefficients $a_n$ may be zero, if $\mu_1, \mu_2, \cdots$ is a proper subsequence of $\lambda_1, \lambda_2, \cdots$ and not identical with the latter. If the coefficient of $z^{\lambda_n}$ in (28) is 0, add the term

$$\frac{z^{\lambda_n}}{\lambda_n!}$$

to (28). The series obtained in this way is exactly of the form (1) and satisfies all requirements of Problem I.

Brown University,
Providence, R. I.

# NÖRLUND SUMMABILITY OF DOUBLE FOURIER SERIES

BY

JOHN G. HERRIOT

1. **Introduction.** Throughout this paper the function $f(t, u)$ is assumed to to be Lebesgue integrable over the square $Q$ $(-\pi, \pi; -\pi, \pi)$ and to have period $2\pi$ in each variable. The double Fourier series of $f$ is denoted by $\sigma(f)$ and the rectangular partial sums of $\sigma(f)$ are denoted by $s_{mn}(x, y; f)$. To say that a method of summability $S$ possesses the localization property means that if $f$ vanishes in a neighborhood of $(x, y)$ then $S$ sums $\sigma(f)$ at $(x, y)$ to 0. It is well known that the Cesàro method $(C, 1, 1)$, for example, does not possess the localization property. G. Grünwald [2](1) has shown that at any point $(x, y)$ of continuity of $f$ the square partial sums $s_{nn}(x, y; f)$ are summable $(C, 1)$ to $f(x, y)$. Thus $(C, 1)$ applied to the square partial sums possesses the localization property. We show in §5 that this is the best possible result.

In this paper we shall apply Nörlund means to $\sigma(f)$. To define the Nörlund mean of $\{s_{nn}(x, y; f)\}$ let $\{p_n\}$ be any sequence of constants. Let $P_n = \sum_{k=0}^{n} p_k \neq 0$. The Nörlund mean is

$$(1.01) \qquad t_n(x, y; f) = \frac{1}{P_n} \sum_{k=0}^{n} p_{n-k} s_{kk}(x, y; f).$$

If $t_n(x, y; f)$ tends to a limit as $n \to \infty$ the sequence $\{s_{nn}(x, y; f)\}$ is said to be summable $N_p$ to this limit. We shall consider only regular Nörlund methods of summability. The conditions of regularity for $N_p$ are(2)

$$(1.02) \qquad \sum_{k=0}^{n} |p_k| = O(|P_n|), \qquad p_n/P_n \to 0 \quad \text{as} \quad n \to \infty.$$

Cesàro $(C, \alpha)$, $\alpha > 0$, is clearly a regular Nörlund method.

We shall also consider a double Nörlund transform of $\{s_{mn}(x, y; f)\}$. Let $\{p_n^{(k)}\}$ $(k=1, 2)$ be two sequences of constants. Let $P_n^{(k)} = \sum_{j=0}^{n} p_j^{(k)} \neq 0$. Then the double Nörlund transform is

$$(1.03) \qquad t_{mn}(x, y; f) = \frac{1}{P_m^{(1)} P_n^{(2)}} \sum_{j,k=0}^{m,n} p_{m-j}^{(1)} p_{n-k}^{(2)} s_{jk}(x, y; f).$$

We shall restrict the manner in which $m, n \to \infty$. If, for any $\lambda \geq 1$, $t_{mn}(x, y; f)$ tends to a limit when $m, n \to \infty$ in such a manner that $m/n \leq \lambda$, $n/m \leq \lambda$, this

Presented to the Society, September 11, 1940; received by the editors May 26, 1941.

(1) The numbers in square brackets refer to the bibliography at the end of the paper.

(2) See, for example, Hille and Tamarkin [4, p. 758].

limit being independent of $\lambda$, then $\sigma(f)$ is said to be restrictedly summable $N_p$ at $(x, y)$ to this limit. $(C, \alpha, \beta)$ is clearly a double Nörlund method.

In §§5 and 6 of this paper local conditions are imposed on the function whose double Fourier series is under consideration in order to discover which of these methods of summability possess the localization property and which do not. In §§7 to 11 methods of summability which sum $\sigma(f)$ almost everywhere to $f$ are studied. Theorem 5 is a generalization of and includes the result of Marcinkiewicz and Zygmund [6]. When the present paper had been prepared for publication the author received a copy of a paper just published by Grünwald [3] in which it was shown that the sequence $\{s_{nn}(x, y; f)\}$ is summable $(C, 1)$ almost everywhere to $f(x, y)$. However, by Corollary 6.1 of the present paper, this result is true also for $(C, \alpha)$, $\alpha > 0$. Both Corollary 6.1 and Theorem 6 from which it follows were established several months before the appearance of Grünwald's paper. Indeed the result of Corollary 6.1 was known much earlier, for, on reading the proofs of a paper of Marcinkiewicz [5] in which it was shown that the sequence $\{s_{nn}(x, y; f)\}$ is summable $(C, 2)$ almost everywhere to $f(x, y)$, Zygmund pointed out that the result could be extended to $(C, \alpha)$, $\alpha > 0$. But Marcinkiewicz did not wish to change his paper and so the result was not published.

2. **Basic formulas.** The following notation will be employed throughout this paper. Let

$$\begin{aligned} \phi_{xy}(t, u) = f(x + t, y + u) + f(x + t, y - u) + f(x - t, y + u) \\ + f(x - t, y - u) - 4f(x, y). \end{aligned} \tag{2.01}$$

It is well known that

$$s_{mn}(x, y; f) = \frac{1}{\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(x + t, y + u)D_m(t)D_n(u)dtdu \tag{2.02}$$

where $D_m(t)$ denotes the Dirichlet kernel. Then

$$t_n(x, y; f) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(x + t, y + u)K_n(t, u)dtdu \tag{2.03}$$

where

$$\begin{aligned} K_n(t, u) &= \frac{1}{\pi^2 P_n}\sum_{k=0}^{n} p_{n-k}D_k(t)D_k(u) \\ &= \frac{1}{\pi^2 P_n}\sum_{k=0}^{n} \frac{p_{n-k}\sin (k + \frac{1}{2})t \sin (k + \frac{1}{2})u}{4 \sin \frac{1}{2}t \sin \frac{1}{2}u}. \end{aligned} \tag{2.04}$$

Clearly $K_n(t, u)$ is an even-even function of $t$ and $u$ and

$$\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} K_n(t, u)dtdu = 1.$$

It follows that

$$(2.05)\qquad t_n(x, y; f) - f(x, y) = \int_0^\pi \int_0^\pi \phi_{xy}(t, u) K_n(t, u) dt du.$$

In order to obtain alternative forms for $K_n(t, u)$ we set

$$(2.06)\qquad \mathfrak{P}_n(t) = \sum_{k=0}^{n} p_k e^{ikt} = \sum_{k=0}^{n} p_k \cos kt + i \sum_{k=0}^{n} p_k \sin kt = \mathfrak{C}_n(t) + i\mathfrak{S}_n(t).$$

Now $\sin (k+\frac{1}{2})t \sin (k+\frac{1}{2})u = -\frac{1}{2}[\cos (k+\frac{1}{2})(t+u) - \cos (k+\frac{1}{2})(t-u)]$ and

$$\begin{aligned}\sum_{k=0}^{n} p_{n-k} \cos (k + \tfrac{1}{2})(t \pm u) &= \sum_{k=0}^{n} p_k \cos (n - k + \tfrac{1}{2})(t \pm u) \\ &= \mathfrak{C}_n(t \pm u) \cos (n + \tfrac{1}{2})(t \pm u) \\ &\quad + \mathfrak{S}_n(t \pm u) \sin (n + \tfrac{1}{2})(t \pm u).\end{aligned}$$

Substituting in (2.04) we have

$$(2.07)\qquad \begin{aligned}K_n(t, u) = &- (8\pi^2 P_n \sin \tfrac{1}{2}t \sin \tfrac{1}{2}u)^{-1} \{\mathfrak{C}_n(t + u) \cos (n + \tfrac{1}{2})(t + u) \\ &+ \mathfrak{S}_n(t + u) \sin (n + \tfrac{1}{2})(t + u) - \mathfrak{C}_n(t - u) \cos (n + \tfrac{1}{2})(t - u) \\ &- \mathfrak{S}_n(t - u) \sin (n + \tfrac{1}{2})(t - u)\}.\end{aligned}$$

If we apply the mean value theorem to this we obtain

$$(2.08)\qquad \begin{aligned}K_n(t, u) = &- u(4\pi^2 P_n \sin \tfrac{1}{2}t \sin \tfrac{1}{2}u)^{-1} \{- \mathfrak{C}_n(\xi_1) \cdot (n + \tfrac{1}{2}) \sin (n + \tfrac{1}{2})\xi_1 \\ &+ \mathfrak{C}_n'(\xi_1) \cos (n + \tfrac{1}{2})\xi_1 + \mathfrak{S}_n(\xi_2) \cdot (n + \tfrac{1}{2}) \cos (n + \tfrac{1}{2})\xi_2 \\ &+ \mathfrak{S}_n'(\xi_2) \sin (n + \tfrac{1}{2})\xi_2\}, \qquad t - u \leqq \xi_1, \xi_2 \leqq t + u.\end{aligned}$$

Forming the double Nörlund transform of $s_{mn}(x, y; f)$ we have

$$(2.09)\qquad t_{mn}(x, y; f) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} f(x + t, y + u) N_m^{(1)}(t) N_n^{(2)}(u) dt du$$

where

$$(2.10)\qquad N_n^{(k)}(t) = \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n} p_{n-j}^{(k)} D_j(t) = \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n} \frac{p_{n-j}^{(k)} \sin (j + \frac{1}{2})t}{2 \sin \frac{1}{2}t}, \qquad k = 1, 2.$$

Thus $N_n^{(k)}(t)$ is an even function of $t$ and

$$\int_{-\pi}^{\pi} N_n^{(k)}(t) dt = 1, \qquad k = 1, 2.$$

We easily deduce that

$$(2.11)\qquad t_{mn}(x, y; f) - f(x, y) = \int_0^\pi \int_0^\pi \phi_{xy}(t, u) N_m^{(1)}(t) N_n^{(2)}(u) dt du.$$

Defining $\mathfrak{P}_n^{(k)}(t)$, $\mathfrak{C}_n^{(k)}(t)$, $\mathfrak{S}_n^{(k)}(t)$ analogously to (2.06) and proceeding as in the deduction of (2.07) we obtain

$$(2.12)\ N_n^{(k)}(t) = (2\pi P_n^{(k)} \sin \tfrac{1}{2}t)^{-1}\{\mathfrak{C}_n^{(k)}(t) \sin (n + \tfrac{1}{2})t - \mathfrak{S}_n^{(k)}(t) \cos (n + \tfrac{1}{2})t\},$$
$$k = 1, 2.$$

3. **Estimates of the kernels.** We require estimates for $K_n(t, u)$ and $N_n^{(k)}(t)$. We shall assume throughout this section that the sequences $\{p_n\}$ and $\{p_n^{(k)}\}$ $(k=1, 2)$ satisfy (1.02) and that $n|p_n| = O(|P_n|)$, $n|p_n^{(k)}| = O(|P_n^{(k)}|)$. All $\{p_n\}$ and $\{p_n^{(k)}\}$ used in our theorems satisfy these conditions.

Since $|D_k(t)| \leqq k+1/2$, it follows from (2.04) that(3)

$$(3.01)\qquad |K_n(t, u)| \leqq An^2, \qquad n \geqq 1, \text{ all } t, u.$$

Also from (2.04) we have

$$(3.02)\qquad |K_n(t, u)| \leqq A/tu, \qquad 0 < t, u \leqq \pi.$$

In the same way from (2.10) we obtain

$$(3.03)\qquad |N_n^{(k)}(t)| \leqq An, \qquad n \geqq 1, \text{ all } t,\ k = 1, 2.$$

In order to obtain further estimates for the kernels we need to estimate $\mathfrak{P}_n(t)$ and $\mathfrak{P}_n'(t)$. We put

$$(3.04)\qquad |p_n| = r_n, \qquad R_n = \sum_{k=0}^{n} r_k, \qquad V_0 = 0, \qquad V_n = \sum_{k=1}^{n} |p_k - p_{k-1}|,$$

and introduce the step functions

$$(3.05)\qquad r(u) = r_{[u]}, \qquad R(u) = R_{[u]}, \qquad V(u) = V_{[u]},$$

where $[u]$ as usual denotes the largest integer less than or equal to $u$. Let us note that by (1.02)

$$(3.06)\qquad |P_n| = \left|\sum_{k=0}^{n} p_k\right| \leqq \sum_{k=0}^{n} |p_k| = R_n = \sum_{k=0}^{n} |p_k| \leqq A|P_n|.$$

Proceeding as on p. 768 of the paper of Hille and Tamarkin [4] and noting that $t^{-1}r(1/t) \leqq AR(1/t)$ we have

$$(3.07)\qquad |\mathfrak{P}_n(t)| \leqq A\left\{R\left(\frac{1}{t}\right) + \frac{1}{t}\left[r_n + V_n - V\left(\frac{1}{t}\right)\right]\right\}, \qquad \frac{1}{n} \leqq t \leqq \frac{3\pi}{2}.$$

If we set

$$(3.08)\qquad U_k' = i\sum_{j=0}^{k} je^{ijt} = \frac{d}{dt}\left\{\sum_{j=0}^{k} e^{ijt}\right\},$$

(3) Here and in the sequel the letter $A$ denotes an absolute constant. The constant need not be the same at every occurrence.

then for $1/n \leqq t \leqq 3\pi/2$ we have $|tU_k'| \leqq An$ ($k=0, 1, 2, \cdots, n$). Using this fact and proceeding as in proving (3.07) we get

$$(3.09) \qquad |\mathfrak{P}_n'(t)| \leqq An\left\{R\left(\frac{1}{t}\right) + \frac{1}{t}\left[r_n + V_n - V\left(\frac{1}{t}\right)\right]\right\}, \qquad \frac{1}{n} \leqq t \leqq \frac{3\pi}{2}.$$

Then for $t, u > 0$, $1/n \leqq t+u \leqq 3\pi/2$, $|t-u| \geqq 1/n$ we have

$$(3.10) \qquad \begin{aligned} |K_n(t, u)| \leqq \frac{A}{R_n tu}&\left\{R\left(\frac{1}{t+u}\right) + \frac{1}{t+u}\left[r_n + V_n - V\left(\frac{1}{t+u}\right)\right]\right. \\ &\left. + R\left(\frac{1}{|t-u|}\right) + \frac{1}{|t-u|}\left[r_n + V_n - V\left(\frac{1}{|t-u|}\right)\right]\right\}, \end{aligned}$$

$$(3.11) \qquad \begin{aligned} |K_n(t, u)| \leqq \frac{An}{R_n(t+u)}&\left\{R\left(\frac{1}{|t-u|}\right)\right. \\ &\left. + \frac{1}{|t-u|}\left[r_n + V_n - V\left(\frac{1}{t+u}\right)\right]\right\}. \end{aligned}$$

Relation (3.10) follows from (2.07) and (3.07); (3.11) follows from (2.08), (3.07) and (3.09) if we note that $K_n(t, u) = K_n(u, t)$ and that $t+u \leqq 2t$ in case $t-u \geqq 1/n$.

Analogously to (3.04) and (3.05) we can define $r_n^{(k)}$, $R_n^{(k)}$, $V_n^{(k)}$, $r^{(k)}(u)$, $R^{(k)}(u)$, $V^{(k)}(u)$ and obtain an estimate for $|\mathfrak{P}_n^{(k)}(t)|$ similar to (3.07). Then from (2.12) we have

$$(3.12) \qquad |N_n^{(k)}(t)| \leqq A\{M_{n1}^{(k)}(t) + M_{n2}^{(k)}(t) + M_{n3}^{(k)}(t)\}, \quad 1/n \leqq t \leqq \pi,\ k = 1, 2,$$

where

$$(3.13) \qquad \begin{aligned} M_{n1}^{(k)}(t) &= \frac{1}{tR_n^{(k)}} R^{(k)}\left(\frac{1}{t}\right), \qquad M_{n2}^{(k)}(t) = \frac{1}{t^2 R_n^{(k)}} r_n^{(k)}, \\ M_{n3}^{(k)}(t) &= \frac{1}{t^2 R_n^{(k)}}\left\{V_n^{(k)} - V^{(k)}\left(\frac{1}{t}\right)\right\}, \qquad k = 1, 2. \end{aligned}$$

Estimating $\mathfrak{P}_n(t)$ and $\mathfrak{P}_n^{(k)}(t)$ as on p. 767 of the paper of Hille and Tamarkin [4] we obtain from (2.07) and (2.12), respectively,

$$(3.14) \qquad \begin{aligned} &|K_n(t, u)| \leqq \{A(\delta)/R_n\}\{V_n + r_n\}, \\ &\qquad\qquad 0 < \delta \leqq t, u \leqq \pi,\ t + u \leqq 2\pi - \delta,\ |t - u| \geqq \delta, \end{aligned}$$

$$(3.15) \qquad |N_n^{(k)}(t)| \leqq \{A(\delta)/R_n^{(k)}\}\{V_n^{(k)} + r_n^{(k)}\}, \qquad 0 < \delta \leqq t \leqq \pi,\ k = 1, 2,$$

where $A(\delta)$ depends only on $\delta$.

Finally we consider the $(C, 1)$ kernel $K_n^1(t, u)$ which is a special case of $K_n(t, u)$ when $p_n \equiv 1$. In case $n \geqq 1$, $0 \leqq t \leqq \pi$, $0 \leqq u \leqq \pi/2$ or $0 \leqq t \leqq \pi/2$,

$\pi/2 \leqq u \leqq \pi$ we shall show that

$$(3.16)\qquad \begin{aligned} |K_n^1(t,u)| \leqq &\frac{An^2}{(1+n^{3/2}t^{3/2})(1+n^{3/2}u^{3/2})} \\ &+ \frac{An^2}{[1+n^{3/2}|t-u|^{3/2}][1+n^{3/2}(t+u)^{3/2}]}, \end{aligned}$$

the positive square root being taken in all cases. Since $p_n \equiv 1$, we have $r_n = 1$, $P_n = R_n = n+1$, $V_n = 0$ ($n = 0, 1, 2, \cdots$). Then, from (3.01), (3.02), (3.10) and (3.11) we have, respectively,

$$(3.17)\qquad |K_n^1(t,u)| \leqq An^2, \qquad n \geqq 1, \text{ all } t, u,$$

$$(3.18)\qquad |K_n^1(t,u)| \leqq A/tu, \qquad 0 < t, u \leqq \pi,$$

$$(3.19)\qquad |K_n^1(t,u)| \leqq \frac{A}{ntu(t+u)} + \frac{A}{ntu|t-u|} \qquad \frac{1}{n} \leqq t+u \leqq \frac{3\pi}{2},$$

$$(3.20)\qquad |K_n^1(t,u)| \leqq \frac{A}{|t-u|(t+u)} \qquad |t-u| \geqq \frac{1}{n}, \quad t, u > 0.$$

Let $D_1$ be the part of the domain under consideration in which $t \leqq 2/n$, $u \leqq 2/n$, $D_2$ that part in which $t > 2/n$, $u \leqq 1/n$, $D_4$ the part in which $0 \leqq t-u \leqq 1/n$, $t > 2/n$, $u > 1/n$, $D_6$ the part in which $t > 2/n$, $1/n < u \leqq t/2$, $D_8$ the part in which $t > 2/n$, $t/2 < u < t - 1/n$, and $D_3$, $D_5$, $D_7$, $D_9$ the domains symmetric to $D_2$, $D_4$, $D_6$, $D_8$, respectively. Then (3.16) follows from (3.17) in $D_1$, from (3.20) in $D_2$, from (3.18) in $D_4$, and from (3.19) in $D_6$ and $D_8$. It follows in $D_3$, $D_5$ $D_7$, $D_9$ by symmetry. Thus (3.16) is completely established.

4. **Preliminary lemmas.** The following lemmas concerning the Nörlund coefficients $p_n$ and $P_n$ will be useful.

LEMMA 1. *If* $\sum_{k=1}^n k|p_k - p_{k-1}| = O(|P_n|)$, *then* $n|p_n| = O(|P_n|)$ *and* $\sum_{k=0}^n |p_k| = O(|P_n|)$.

LEMMA 2. *If* $n\sum_{k=1}^n |p_k - p_{k-1}| = O(|P_n|)$, *then* $n = O(|P_n|)$ *and* $\sum_{k=1}^n |P_k|/k = O(|P_n|)$.

It is clear that the hypothesis of Lemma 2 implies that of Lemma 1. These lemmas follow easily from the relations

$$P_k = (k+1)p_k + \sum_{j=1}^{k} j(p_{j-1} - p_j),$$

$$p_k = p_n + \sum_{j=k+1}^{n} (p_{j-1} - p_j), \qquad k = 0, 1, 2, \cdots, n-1.$$

We may also easily establish the following analogue of Abel's partial sum formula.

LEMMA 3. *Let* $\{a_{jk}\}$, $\{b_{jk}\}$ *be two sequences. Let*

$$\Delta_{10}a_{jk} = a_{jk} - a_{j+1,k}, \qquad \Delta_{01}a_{jk} = a_{jk} - a_{j,k+1}, \qquad \Delta_{11}a_{jk} = \Delta_{01}\Delta_{10}a_{jk}.$$

*Similarly define* $\Delta_{10}b_{jk}$, $\Delta_{01}b_{jk}$, $\Delta_{11}b_{jk}$. *Then*

$$\begin{aligned}\sum_{j=c,k=d}^{m,n} a_{jk}\Delta_{11}b_{jk} = {} & \sum_{j=c,k=d}^{m,n} b_{jk}\Delta_{11}a_{j-1,k-1} - \sum_{j=c}^{m} b_{jd}\Delta_{10}a_{j-1,d-1} \\ & + \sum_{j=c}^{m} b_{j,n+1}\Delta_{10}a_{j-1,n} - \sum_{k=d}^{n} b_{ck}\Delta_{01}a_{c-1,k-1} \\ & + \sum_{k=d}^{n} b_{m+1,k}\Delta_{01}a_{m,k-1} + a_{c-1,d-1}b_{cd} \\ & - a_{m,d-1}b_{m+1,d} - a_{c-1,n}b_{c,n+1} + a_{mn}b_{m+1,n+1}.\end{aligned} \tag{4.01}$$

5. **Local results making use of square partial sums.** Our first theorem extends the result of Grünwald [2] in two directions and also includes his result.

THEOREM 1. *Let* $N_p$ *be a regular Nörlund method of summability satisfying the condition*

$$\sum_{k=1}^{n-1} (n-k) \mid p_k - p_{k-1} \mid = O(\mid P_n \mid). \tag{5.01}$$

*Then at any point* $(x, y)$ *such that*

$$\begin{aligned}\Phi(h, k) &= \int_0^h dt \int_0^k \mid \phi_{xy}(t, u) \mid du = o(hk), \\ \Phi^*(h, k) &= \int_0^h dt \int_0^k \left| \phi_{xy}\left(\frac{t-u}{2^{1/2}}, \frac{t+u}{2^{1/2}}\right) \right| du = o(hk)\end{aligned} \tag{5.02}$$

*as* $h, k \to 0$ *simultaneously but independently, the sequence* $\{s_{nn}(x, y; f)\}$ *is summable* $N_p$ *to* $f(x, y)$.

It should be noted that the second condition on the function at $(x, y)$ is similar to the first. The first is applied to rectangles along the axes, the second to rectangles along the bisectors of the angles between the axes. The factors $2^{-1/2}$ are not essential, but are introduced for convenience.

**Proof.** A regular Nörlund method $N_p$ includes $(C, 1)$ if[4]

$$n \mid p_0 \mid + \sum_{k=1}^{n-1} (n-k) \mid p_k - p_{k-1} \mid < A \mid P_n \mid. \tag{5.03}$$

Hence if $N_p$ satisfies (5.01) and is regular, then it includes $(C, 1)$. Thus it

---

[4] See, for example, Hille and Tamarkin [4, p. 782].

suffices to prove the theorem for $(C, 1)$. Let $t_n^1(x, y; f)$ denote the $(C, 1)$ transform of the sequence $\{s_{nn}(x, y; f)\}$. From (2.05) we have

$$t_n^1(x, y; f) - f(x, y) = \int_0^\pi \int_0^\pi \phi_{xy}(t, u) K_n^1(t, u)\,dtdu \tag{5.04}$$

where $K_n^1(t, u)$ is given by (2.04) with $p_n \equiv 1$. Fix $(x, y)$.

Given $\epsilon > 0$ we can choose $\delta$ such that $0 < \delta < \pi/4$ and such that

$$|\Phi(h, k)| < \epsilon |hk|, \qquad |\Phi^*(h, k)| < \epsilon |hk|, \text{ for } 0 < |h|, \ |k| \leqq 2\delta. \tag{5.05}$$

Suppose $n > 2/\delta$. Let $B_\delta = [0, \pi; 0, \pi] - [0, \delta; 0, \delta]$. Then

$$\begin{aligned} |t_n^1(x, y; f) - f(x, y)| &\leqq \left( \int_0^\delta \int_0^\delta + \iint_{B_\delta} \right) \cdot |\phi_{xy}(t, u) K_n^1(t, u)|\,dtdu \\ &= J_1 + J_2. \end{aligned} \tag{5.06}$$

Then by (3.16)

$$\begin{aligned} J_1 \leqq\ & An^2 \int_0^\delta \int_0^\delta \frac{|\phi_{xy}(t, u)|\,dtdu}{(1 + n^{3/2}t^{3/2})(1 + n^{3/2}u^{3/2})} \\ & + An^2 \int_0^\delta \int_0^\delta \frac{|\phi_{xy}(t, u)|\,dtdu}{[1 + n^{3/2}|t - u|^{3/2}][1 + n^{3/2}(t + u)^{3/2}]} = J_{11} + J_{12}. \end{aligned}$$

Integrating $J_{11}$ twice by parts and applying (5.05) we get

$$\begin{aligned} J_{11} \leqq\ & \frac{An^2\epsilon\delta^2}{n^3\delta^3} + \frac{An^{7/2}\delta\epsilon}{n^{3/2}\delta^{3/2}} \int_0^\delta \frac{u^{3/2}du}{(1 + n^{3/2}u^{3/2})^2} \\ & + An^5\epsilon \int_0^\delta \int_0^\delta \frac{t^{3/2}u^{3/2}dtdu}{(1 + n^{3/2}t^{3/2})^2(1 + n^{3/2}u^{3/2})^2}. \end{aligned}$$

But

$$\int_0^\delta \frac{u^{3/2}du}{(1 + n^{3/2}u^{3/2})^2} = \int_0^{1/n} + \int_{1/n}^\delta \leqq \frac{1}{n} \cdot \frac{1}{n^{3/2}} + \frac{1}{n^3} \int_{1/n}^\delta \frac{du}{u^{3/2}} \leqq \frac{3}{n^{5/2}}.$$

Hence, since $n\delta > 2$, we easily obtain $J_{11} \leqq A\epsilon$. Applying the transformation $t = 2^{-1/2}(t' - u')$, $u = 2^{-1/2}(t' + u')$ to $J_{12}$ and proceeding as above we get $J_{12} \leqq A\epsilon$. Thus $J_1 \leqq A\epsilon$.

Next let $B_1$ be that part of $B_\delta$ in which $t \geqq \delta$, $u \leqq \delta' < \delta/4$, $B_2$ the domain symmetric to $B_1$, $B_3$ that part of $B_\delta$ in which $|t - u| \leqq \delta'$, $B_4$ the rest of $B_\delta$. Clearly $B_1 \cdot B_3 = B_2 \cdot B_3 = 0$ since $\delta' < \delta/4$. In $B_1 + B_2 + B_3$ we have $|K_n^1(t, u)| \leqq A/(3\delta/4)^2$. This follows from (3.20) in $B_1 + B_2$ and from (3.18) in $B_3$. Since $\phi_{xy}(t, u)$ is integrable we can choose $\delta'$ depending only on $\delta$ (and hence on $\epsilon$), $0 < \delta' < \delta/4$, such that

$$\iint_{B_1 + B_2 + B_3} |\phi_{xy}(t, u) K_n^1(t, u)|\,dtdu < \epsilon. \tag{5.07}$$

Fixing $\delta'$, we see that, on account of (3.14), $K_n^1(t, u)\to 0$ uniformly in $B_4$. Thus for all sufficiently large $n$ we have $J_2 < 2\epsilon$ and consequently $|t_n^1(x, y; f) - f(x, y)| \leqq A\epsilon$. That is, $t_n^1(x, y; f)\to f(x, y)$ as $n\to\infty$. This completes the proof of the theorem.

COROLLARY 1.1. *Let $N_p$ be a regular Nörlund method of summability satisfying* (5.01). *Then $N_p$ applied to the square partial sums of the double Fourier series possesses the localization property.*

For if $f$ vanishes in a neighborhood of $(x, y)$, $\phi_{xy}(t, u)$ satisfies (5.02).

Before showing that (5.01) is also partly necessary in order that $N_p$ applied to the square partial sums should possess the localization property we prove the following lemma.

LEMMA 4. *Let $N_p$ be a regular Nörlund method of summability with $p_n \geqq 0$, $p_n$ non-increasing, $p_1 < p_0$, $n/P_n\to\infty$ as $n\to\infty$. Suppose $0<\delta<\pi$. Let $E = [-\pi, \pi; -\pi, \pi] - (-\delta, \delta; -\delta, \delta)$. Then there exists $N>0$ such that*

$$\text{(5.08)} \qquad \operatorname*{ess\,sup}_{(t,u)\in E} |K_n(t, u)| > An/P_n, \qquad \textit{all } n > N, A > 0.$$

**Proof.** From (2.04) we have

$$\begin{aligned} K_n(\pi, 0) &= \frac{1}{\pi^2 P_n}\sum_{k=0}^{n} p_{n-k}\cdot\tfrac{1}{2}(-1)^k(k+\tfrac{1}{2}) \\ &= \frac{(-1)^n}{2\pi^2 P_n}(n+\tfrac{1}{2})\sum_{k=0}^{n}(-1)^k p_k - \frac{(-1)^n}{2\pi^2 P_n}\sum_{k=0}^{n}(-1)^k k p_k \\ &= J_1 - J_2. \end{aligned}$$

Since $p_n$ is non-increasing we have immediately $|J_1| \geqq n(p_0 - p_1)/2\pi^2 P_n$. If we set $W_k = \sum_{j=0}^{k}(-1)^j j$, then $|W_k| \leqq k$ and we easily get

$$\begin{aligned} \left|\sum_{k=0}^{n}(-1)^k k p_k\right| &= \left|\sum_{k=0}^{n-1} W_k(p_k - p_{k+1}) + W_n p_n\right| \leqq \sum_{k=0}^{n-1} k(p_k - p_{k+1}) + np_n \\ &= \sum_{k=1}^{n} p_k \leqq P_n. \end{aligned}$$

Hence $|J_2| \leqq 1/2\pi^2$. But we can choose $N>0$ such that $n/P_n > 2/(p_0 - p_1)$ for all $n > N$. Then for $n > N$ we find $|K_n(\pi, 0)| \geqq |J_1| - |J_2| \geqq n(p_0 - p_1)/4\pi^2 P_n = An/P_n$, $A > 0$. But $E$ is closed and for each $n$, $K_n(t, u)$ is continuous. Thus (5.08) follows.

THEOREM 2. *Let $N_p$ be a regular Nörlund method of summability with $p_n \geqq 0$, $p_n$ non-increasing, $p_1 < p_0$, $n/P_n\to\infty$ as $n\to\infty$. Then there exists $f$ vanishing in a neighborhood of* (0, 0) *such that* $\limsup_{n\to\infty} |t_n(0, 0; f)| = +\infty$.

**Proof.** Let $E=[-\pi, \pi; -\pi, \pi]-(-\delta, \delta; -\delta, \delta)$, $0<\delta<\pi$. Consider the class of functions $f\in L[-\pi, \pi; -\pi, \pi]$ which vanish in $(-\delta, \delta; -\delta, \delta)$, that is, the class of functions $f\in L(E)$. Then

$$t_n(0, 0; f) = \iint_E f(t, u)K_n(t, u)dtdu = T_n(f) \tag{5.09}$$

defines a linear functional on the space $L(E)$ with norm

$$\|T_n\| = \operatorname*{ess\,sup}_{(t, u)\in E} | K_n(t, u) |. \tag{5.10}$$

Now suppose that the conclusion of our theorem does not hold. Then for every $f\in L(E)$, $\limsup_{n\to\infty} | T_n(f)| < \infty$. By a well known theorem of Banach and Steinhaus([5]) it follows that $\|T_n\| \leqq M < \infty$ for all $n$. Thus by (5.10) and (5.08) we have $An/P_n \leqq M$, $A>0$, for all $n>N$. This contradicts the hypothesis.

COROLLARY 2.1. *Let $N_p$ be a regular Nörlund method of summability with $p_n\geqq 0$, $p_n$ non-increasing, $p_1<p_0$. Then* (5.01) *is necessary as well as sufficient for $N_p$ applied to the square partial sums of the double Fourier series to possess the localization property.*

**Proof.** To prove the necessity we first note that $n/P_n$ is non-decreasing since $p_n$ is non-increasing. Then in order that $N_p$ applied to the square partial sums should possess the localization property we must have $n/P_n$ bounded by Theorem 2. The condition (5.01) is an immediate consequence of this.

The case in which $p_n\geqq 0$, $p_n$ non-increasing, is especially important as it includes Cesàro $(C, \alpha)$, $0<\alpha\leqq 1$. Because of the simplicity of the result in this case we state it separately.

COROLLARY 2.2. *Under the hypotheses of Corollary* 2.1, *a necessary and sufficient condition that $N_p$ applied to the square partial sums of the double Fourier series should possess the localization property is $n=O(P_n)$.*

From this it follows that $(C, \alpha)$ applied to the square partial sums possesses the localization property if and only if $\alpha\geqq 1$. Thus Grünwald's [2] result is the best possible in the sense that it cannot be extended to $(C, \alpha)$, $\alpha<1$.

6. **Local properties of restricted summability.** We turn now to restricted double Nörlund summability of the rectangular partial sums of the double Fourier series. The results are similar to those in §5.

THEOREM 3. *Let $N_p$ be a double Nörlund method of summability satisfying the conditions*

$$n\cdot\sum_{j=1}^{n} | p_j^{(k)} - p_{j-1}^{(k)} | = O(| P_n^{(k)} |), \qquad k = 1, 2. \tag{6.01}$$

---

([5]) See, for example, Banach [1, p. 80].

*Then at any point* $(x, y)$ *such that*

$$\Phi(h, k) = \int_0^h dt \int_0^k | \phi_{xy}(t, u) | du = o(hk) \tag{6.02}$$

as $h, k \to 0$ *simultaneously but independently,* $\sigma(f)$ *is restrictedly summable* $N_p$ *to* $f(x, y)$.

**Proof.** Let $\lambda \geqq 1$ be any fixed number. It suffices to show that $t_{mn}(x, y; f) \to f(x, y)$ as $m, n \to \infty$ in such a manner that $m/n \leqq \lambda$, $n/m \leqq \lambda$. Fix $(x, y)$. Given $\epsilon > 0$, we can choose $\delta > 0$ such that $1/\delta$ is an integer greater than 2 and such that

$$\Phi(h, k) < \epsilon hk, \tag{6.03}$$

whenever $0 < h, k \leqq \delta$. Then from (2.11) we obtain

$$\begin{aligned} | t_{mn}(x, y; f) - f(x, y) | &\leqq \Big( \int_\delta^\pi \int_\delta^\pi + \int_0^\delta \int_\delta^\pi + \int_\delta^\pi \int_0^\delta \\ &\quad + \int_0^\delta \int_0^\delta \Big) | \phi_{xy}(t, u) N_m^{(1)}(t) N_n^{(2)}(u) | dudt \\ &= J_1 + J_2 + J_3 + J_4. \end{aligned} \tag{6.04}$$

On account of Lemma 1, the estimates of §3 may be applied here. From (3.15), (6.01) and Lemma 1 we have that $N_m^{(1)}(t)N_n^{(2)}(u) \to 0$ uniformly in $[\delta, \pi; \delta, \pi]$. Thus $J_1$ is small for all sufficiently large $m$ and $n$. If $m/n \leqq \lambda$, $n/m \leqq \lambda$ we see from (3.03), (3.15), (6.01) and Lemma 1 that $N_m^{(1)}(t)N_n^{(2)}(u)$ is bounded in the domains of integration of $J_2$ and $J_3$. Thus we can find $\delta'$ such that $0 < \delta' < \delta$ and so small that

$$\Big( \int_0^{\delta'} \int_\delta^\pi + \int_\delta^\pi \int_0^{\delta'} \Big) | \phi_{xy}(t, u) N_m^{(1)}(t) N_n^{(2)}(u) | dudt$$

is uniformly small. In the remainder of the domains of integration of $J_2$ and $J_3$, $N_m^{(1)}(t)N_n^{(2)}(u) \to 0$ uniformly and thus $J_2 + J_3$ is small for all sufficiently large $m$ and $n$ such that $m/n \leqq \lambda$, $n/m \leqq \lambda$. Thus we can find $N_0 > 1/\delta$ such that $J_1 + J_2 + J_3 < A\epsilon$ if $m, n > N_0$, $m/n \leqq \lambda$, $n/m \leqq \lambda$. In the following we suppose $m, n > N_0$. Then

$$\begin{aligned} J_4 = \Big( &\int_0^{1/m} \int_0^{1/n} + \int_{1/m}^\delta \int_0^{1/n} + \int_0^{1/m} \int_{1/n}^\delta \\ &+ \int_{1/m}^\delta \int_{1/n}^\delta \Big) | \phi_{xy}(t, u) N_m^{(1)}(t) N_n^{(2)}(u) | dudt \\ = J_{41} &+ J_{42} + J_{43} + J_{44}. \end{aligned} \tag{6.05}$$

Then from (3.03) and (6.03) we have at once $J_{41} \leqq A\epsilon$. Also by (3.03) and (3.12) we have

$$J_{42} \leqq An \int_0^{1/n} du \int_{1/m}^{\delta} |\phi_{xy}(t, u)| \left\{M_{m1}^{(1)}(t) + M_{m2}^{(1)}(t) + M_{m3}^{(1)}(t)\right\} dt \tag{6.06}$$
$$= J_{42}^1 + J_{42}^2 + J_{42}^3.$$

Then from (3.13)

$$J_{42}^1 = An \sum_{j=1/\delta}^{m-1} \int_0^{1/n} du \int_{1/(j+1)}^{1/j} |\phi_{xy}(t, u)| \frac{1}{tR_m^{(1)}} R^{(1)}\left(\frac{1}{t}\right) dt$$
$$\leqq \frac{An}{R_m^{(1)}} \sum_{j=1/\delta}^{m-1} (j+1) R_{j+1}^{(1)} \Delta_{10} \Phi\left(\frac{1}{j}, \frac{1}{n}\right)$$
$$= \frac{An}{R_m^{(1)}} \left\{ \sum_{j=1/\delta}^{m-1} \Phi\left(\frac{1}{j}, \frac{1}{n}\right) [(j+1)R_{j+1}^{(1)} - jR_j^{(1)}] \right.$$
$$\left. + \frac{1}{\delta} R_{1/\delta}^{(1)} \Phi\left(\delta, \frac{1}{n}\right) - mR_m^{(1)} \Phi\left(\frac{1}{m}, \frac{1}{n}\right) \right\}$$
$$\leqq \frac{A\epsilon}{R_m^{(1)}} \left\{ \sum_{j=1/\delta}^{m-1} \frac{R_{j+1}^{(1)} + jr_{j+1}^{(1)}}{j} + R_{1/\delta}^{(1)} \right\}.$$

But by Lemma 2 and (3.06) we get

$$\sum_{j=1/\delta}^{m-1} \frac{R_{j+1}^{(k)} + jr_{j+1}^{(k)}}{j} \leqq \sum_{j=1/\delta}^{m-1} r_{j+1}^{(k)} + 2 \sum_{j=1/\delta}^{m-1} \frac{R_{j+1}^{(k)}}{j+1} \leqq AR_m^{(k)}, \qquad k = 1, 2. \tag{6.07}$$

Thus $J_{42}^1 \leqq A\epsilon$. Substituting from (3.13) in $J_{42}^2$, integrating twice by parts and using Lemma 1 we have $J_{42}^2 \leqq A\epsilon$. Again from (3.13) we have

$$J_{42}^3 = An \sum_{j=1/\delta}^{m-1} \int_0^{1/n} du \int_{1/(j+1)}^{1/j} |\phi_{xy}(t, u)| \frac{1}{t^2 R_m^{(1)}} \left[ V_m^{(1)} - V^{(1)}\left(\frac{1}{t}\right) \right] dt$$
$$\leqq \frac{An}{R_m^{(1)}} \sum_{j=1/\delta}^{m-1} (j+1)^2 (V_m^{(1)} - V_j^{(1)}) \Delta_{10} \Phi\left(\frac{1}{j}, \frac{1}{n}\right)$$
$$= \frac{An}{R_m^{(1)}} \left\{ \sum_{j=1/\delta}^{m-1} \Phi\left(\frac{1}{j}, \frac{1}{n}\right) [(2j+1)(V_m^{(1)} - V_j^{(1)}) - j^2 |p_j^{(1)} - p_{j-1}^{(1)}|] \right.$$
$$\left. + \frac{1}{\delta^2} (V_m^{(1)} - V_{(1/\delta)-1}^{(1)}) \Phi\left(\delta, \frac{1}{n}\right) - m^2 (V_m^{(1)} - V_{m-1}^{(1)}) \Phi\left(\frac{1}{m}, \frac{1}{n}\right) \right\}$$
$$\leqq \frac{A\epsilon}{R_m^{(1)}} \left\{ \sum_{j=1/\delta}^{m-1} \frac{2j+1}{j} (V_m^{(1)} - V_j^{(1)}) + \frac{1}{\delta} (V_m^{(1)} - V_{(1/\delta)-1}^{(1)}) \right\}.$$

But by (3.06) and (6.01)

$$
\begin{aligned}
\sum_{j=1/\delta}^{m-1} \frac{2j+1}{j} (V_m^{(k)} - V_j^{(k)}) &\leqq 3 \sum_{j=1/\delta}^{m-1} \sum_{s=j+1}^{m} | p_s^{(k)} - p_{s-1}^{(k)} | \\
&= 3 \sum_{s=(1/\delta)+1}^{m} \sum_{j=1/\delta}^{s-1} | p_s^{(k)} - p_{s-1}^{(k)} | \\
&\leqq 3 \sum_{s=(1/\delta)+1}^{m} s | p_s^{(k)} - p_{s-1}^{(k)} | \leqq A R_m^{(k)}, \quad k = 1, 2.
\end{aligned} \tag{6.08}
$$

Likewise

$$
\frac{1}{\delta} (V_m^{(k)} - V_{(1/\delta)-1}^{(k)}) \leqq m \sum_{j=1}^{m} | p_j^{(k)} - p_{j-1}^{(k)} | \leqq A R_m^{(k)}, \qquad k = 1, 2. \tag{6.09}
$$

Thus $J_{42}^3 \leqq A\epsilon$. Substituting in (6.06) we have $J_{42} \leqq A\epsilon$. In the same way $J_{43} \leqq A\epsilon$. Turning now to $J_{44}$ we have by (3.12)

$$
\begin{aligned}
J_{44} &\leqq A \int_{1/m}^{\delta} dt \int_{1/n}^{\delta} | \phi_{xy}(t, u) | \left\{ M_{m1}^{(1)}(t) + M_{m2}^{(1)}(t) + M_{m3}^{(1)}(t) \right\} \\
&\qquad \cdot \left\{ M_{n1}^{(2)}(u) + M_{n2}^{(2)}(u) + M_{n3}^{(2)}(u) \right\} du \\
&= \sum_{j=1}^{9} J_{44}^{j}.
\end{aligned} \tag{6.10}
$$

We now show that $J_{44}^j \leqq A\epsilon$ ($j = 1, 2, 3, \cdots, 9$) as was done with the $J_{42}^j$. For example, let us take $J_{44}^6$. Then from (3.13)

$$
\begin{aligned}
J_{44}^6 &= A \sum_{j,k=1/\delta}^{m-1,n-1} \int_{1/(j+1)}^{1/j} dt \int_{1/(k+1)}^{1/k} | \phi_{xy}(t, u) | \frac{r_m^{(1)}}{t^2 u^2 R_m^{(1)} R_n^{(2)}} \left[ V_n^{(2)} - V^{(2)}\left(\frac{1}{u}\right) \right] du \\
&\leqq \frac{A r_m^{(1)}}{R_m^{(1)} R_n^{(2)}} \sum_{j,k=1/\delta}^{m-1,n-1} (j+1)^2 (k+1)^2 (V_n^{(2)} - V_k^{(2)}) \Delta_{11} \Phi\left(\frac{1}{j}, \frac{1}{k}\right).
\end{aligned}
$$

Applying Lemma 3 and dropping clearly negative terms we obtain

$$
\begin{aligned}
J_{44}^6 \leqq \frac{A r_m^{(1)}}{R_m^{(1)} R_n^{(2)}} &\left\{ \sum_{j,k=1/\delta}^{m-1,n-1} \Phi\left(\frac{1}{j}, \frac{1}{k}\right) \right. \\
&\cdot (2j+1) \left[ (2k+1)(V_n^{(2)} - V_k^{(2)}) - k^2 | p_k^{(2)} - p_{k-1}^{(2)} | \right] \\
&+ \sum_{j=1/\delta}^{m-1} \Phi\left(\frac{1}{j}, \delta\right) \cdot (2j+1) \frac{1}{\delta^2} (V_n^{(2)} - V_{(1/\delta)-1}^{(2)}) \\
&+ \sum_{k=1/\delta}^{n-1} \left[ \frac{1}{\delta^2} \Phi\left(\delta, \frac{1}{k}\right) - m^2 \Phi\left(\frac{1}{m}, \frac{1}{k}\right) \right] \\
&\cdot \left[ (2k+1)(V_n^{(2)} - V_k^{(2)}) - k^2 | p_k^{(2)} - p_{k-1}^{(2)} | \right] \\
&\left. + \frac{1}{\delta^4} (V_n^{(2)} - V_{(1/\delta)-1}^{(2)}) \Phi(\delta, \delta) + m^2 n^2 (V_n^{(2)} - V_{n-1}^{(2)}) \Phi\left(\frac{1}{m}, \frac{1}{n}\right) \right\}.
\end{aligned}
$$

Again dropping those terms which are negative and applying (6.03), (6.01), (6.08), (6.09) and Lemma 1 we obtain $J_{44}^{6} \leqq A\epsilon$. Altogether, then, $J_{44} \leqq A\epsilon$. Combining our estimates in (6.05) and (6.04) we have $|t_{mn}(x, y; f) - f(x, y)| \leqq A\epsilon$ if $m, n > N_0$, $m/n \leqq \lambda$, $n/m \leqq \lambda$. This completes the proof of the theorem.

COROLLARY 3.1. *Let $N_p$ be a double Nörlund method of summability satisfying* (6.01). *Then restricted $N_p$ summability possesses the localization property.*

Before showing (6.01) is also partly necessary in order that restricted $N_p$ possess the localization property we prove the following lemma.

LEMMA 5. *Let $N_p$ be a double Nörlund method of summability with $p_n^{(k)} \geqq 0$, $p_n^{(k)}$ non-increasing* ($k = 1, 2$). *Then*

$$(6.11) \qquad |N_n^{(k)}(\pi)| \geqq (p_0^{(k)} - p_1^{(k)})/2\pi P_n^{(k)}, \quad |N_n^{(k)}(0)| \geqq n/2\pi, \quad k = 1, 2.$$

**Proof.** From (2.10) we have

$$N_n^{(k)}(\pi) = \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n} p_{n-j}^{(k)} \cdot \tfrac{1}{2}(-1)^j = \frac{(-1)^n}{2\pi P_n^{(k)}} \sum_{j=0}^{n} (-1)^j p_j^{(k)}.$$

The first inequality of (6.11) follows immediately. Also from (2.10)

$$\begin{aligned} N_n^{(k)}(0) &= \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n} p_{n-j}^{(k)}(j + \tfrac{1}{2}) = \frac{1}{2\pi} + \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n} (n - j) p_j^{(k)} \\ &= \frac{1}{2\pi} + \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n-1} P_j^{(k)}. \end{aligned}$$

Thus

$$(6.12) \qquad |N_n^{(k)}(0)| \geqq \frac{1}{\pi P_n^{(k)}} \sum_{j=0}^{n-1} P_j^{(k)}.$$

But since $P_n^{(k)}$ is non-decreasing we have $(P_n^{(k)})^{-1} \sum_{j=0}^{n-1} P_j^{(k)} \leqq n$ and thus

$$\begin{aligned} 0 \leqq n - \frac{1}{P_n^{(k)}} \sum_{j=0}^{n-1} P_j^{(k)} &= \frac{1}{P_n^{(k)}} \sum_{j=0}^{n-1} (P_n^{(k)} - P_j^{(k)}) = \frac{1}{P_n^{(k)}} \sum_{j=0}^{n-1} \sum_{s=j+1}^{n} p_s^{(k)} \\ &= \frac{1}{P_n^{(k)}} \sum_{s=1}^{n} s p_s^{(k)} \leqq \frac{1}{P_n^{(k)}} \sum_{s=1}^{n} P_{s-1}^{(k)} = \frac{1}{P_n^{(k)}} \sum_{j=0}^{n-1} P_j^{(k)}, \end{aligned}$$

or

$$n \leqq \frac{2}{P_n^{(k)}} \sum_{j=0}^{n-1} P_j^{(k)}.$$

Substituting this in (6.12) we obtain the second inequality of (6.11).

THEOREM 4. *Let $N_p$ be a double Nörlund method of summability with $p_n^{(k)} \geqq 0$,*

$p_n^{(k)}$ *non-increasing* ($k=1, 2$). *Suppose* $p_1^{(k)} < p_0^{(k)}$, $n/P_n^{(k)} \to \infty$ *as* $n \to \infty$ *for* $k=1$ *or* 2 *or both. Then there exists f vanishing in a neighborhood of* (0, 0) *such that* $\lim \sup_{n\to\infty} |t_{nn}(0, 0; f)| = +\infty$.

The proof is analogous to that of Theorem 2, using Lemma 5 instead of Lemma 4.

As in §5 we may prove the following corollaries:

COROLLARY 4.1. *Let* $N_p$ *be a double Nörlund method of summability with* $p_n^{(k)} \geqq 0$, $p_n^{(k)}$ *non-increasing,* $p_1^{(k)} < p_0^{(k)}$ ($k=1, 2$). *Then* (6.01) *is necessary as well as sufficient in order that restricted* $N_p$ *possess the localization property.*

COROLLARY 4.2. *Under the hypotheses of Corollary* 4.1, *necessary and sufficient conditions that restricted* $N_p$ *summability should possess the localization property are* $n = O(P_n^{(k)})$ ($k=1, 2$).

It follows that restricted $(C, \alpha, \beta)$ possesses the localization property if and only if $\alpha \geqq 1, \beta \geqq 1$.

7. **Preliminary lemmas for almost everywhere results.** We turn now to the study of methods of summability which sum the double Fourier series almost everywhere. The results are generalizations and extensions of those due to Marcinkiewicz and Zygmund [5, 6] and Grünwald [3]. The proofs are based on those given by Marcinkiewicz and Zygmund. We shall require the following lemmas.

LEMMA 6. *Let* $\alpha$ *be any fixed positive number. For* $(x, y)$ *belonging to the square* $Q\ [-\pi, \pi; -\pi, \pi]$, *we write*

$$(7.01) \qquad f_\alpha^*(x, y) = \sup_h \frac{1}{4\alpha h^2} \int_{-\alpha h}^{\alpha h} du \int_{-h}^{h} |f(x+t, y+u)|\, dt,$$

$$(7.02) \qquad f_\alpha^{**}(x, y) = \sup_h \frac{1}{4\alpha h^2} \int_{-\alpha h}^{\alpha h} du \int_{-h}^{h} \left| f\left(x + \frac{t-u}{2^{1/2}},\ y + \frac{t+u}{2^{1/2}}\right) \right| dt$$

*where the number h is so small that the rectangles over which the integrals are taken are contained in* $Q'\ [-2\pi, 2\pi; -2\pi, 2\pi]$. *Let*

$$\mathcal{E}_\alpha^*(\xi) = \underset{(x,y)}{E}\, [f_\alpha^*(x, y) > \xi], \qquad \mathcal{E}_\alpha^{**}(\xi) = \underset{(x,y)}{E}\, [f_\alpha^{**}(x, y) > \xi]$$

*for any* $\xi > 0$. *Then*

$$(7.03) \qquad \begin{matrix} |\mathcal{E}_\alpha^*(\xi)| \\ |\mathcal{E}_\alpha^{**}(\xi)| \end{matrix} \leqq \frac{A}{\xi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |f(x, y)|\, dx dy.$$

In the case of $f_\alpha^*(x, y)$ the proof was given by Marcinkiewicz and Zygmund [6]. For the case of $f_\alpha^{**}(x, y)$ the proof can be carried through in the same way.

LEMMA 7. *Let a be any fixed positive number. For $(x, y)$ belonging to the square $Q$ $[-\pi, \pi; -\pi, \pi]$, we write*

$$f^{*a}(x, y) = \sup_s \left(f_s^{*}(x, y) \cdot 2^{-a|s|}\right), \qquad \text{for } s = 0, \pm 1, \pm 2, \cdots, \tag{7.04}$$

$$f^{**a}(x, y) = \sup_s \left(f_s^{**}(x, y) \cdot 2^{-a|s|}\right), \qquad \text{for } s = 0, \pm 1, \pm 2, \cdots. \tag{7.05}$$

*We write*

$$\mathcal{E}^{*a}(\xi) = \underset{(x,y)}{E} \left[f^{*a}(x, y) > \xi\right], \qquad \mathcal{E}^{**a}(\xi) = \underset{(x,y)}{E} \left[f^{**a}(x, y) > \xi\right]$$

*for any $\xi > 0$. Then*

$$\begin{matrix} |\mathcal{E}^{*a}(\xi)| \\ |\mathcal{E}^{**a}(\xi)| \end{matrix} \leqq \frac{A(a)}{\xi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |f(x, y)| \, dxdy, \tag{7.06}$$

*where $A(a)$ depends only on $a$.*

The proof is similar to that given by Marcinkiewicz and Zygmund for their Lemma 3 [6].

LEMMA 8. *Suppose $P_n \geqq 0$, $P_n$ non-decreasing, $a \geqq 0$. Then the condition*

$$\sum_{k=1}^{n} \frac{P_k}{k} \left(\frac{n}{k}\right)^a = O(P_n) \tag{7.07}$$

*is equivalent to the condition*

$$\sum_{k=0}^{n-1} P_{2^k} \cdot 2^{a(n-k)} = O(P_{2^n}). \tag{7.08}$$

**Proof.** Suppose (7.08) is satisfied. Let $j$ be an integer such that $2^j \leqq n < 2^{j+1}$. Then

$$\begin{aligned} \sum_{k=1}^{n} \frac{P_k}{k} \left(\frac{n}{k}\right)^a &= \sum_{k=0}^{j-1} \left\{ \frac{P_{2^k}}{2^k} \left(\frac{n}{2^k}\right)^a + \frac{P_{2^k+1}}{2^k + 1} \left(\frac{n}{2^k + 1}\right)^a + \cdots \right. \\ &\qquad \left. + \frac{P_{2^{k+1}-1}}{2^{k+1} - 1} \left(\frac{n}{2^{k+1} - 1}\right)^a \right\} + \sum_{k=2^j}^{n} \frac{P_k}{k} \left(\frac{n}{k}\right)^a \\ &\leqq \sum_{k=0}^{j-1} P_{2^{k+1}} \left(\frac{2^{j+1}}{2^k}\right)^a + P_n \cdot 2^{-j} \cdot 2^j \cdot 2^a \\ &\leqq 2^{2a} \sum_{k=0}^{j-1} P_{2^k} \cdot 2^{a(j-k)} + 2^{2a} P_{2^j} + O(P_n) \\ &= O(P_{2^j}) + O(P_n) = O(P_n) \end{aligned}$$

showing (7.07) to be satisfied. The proof of the converse is similar.

8. **Lemma for restricted Nörlund summability.** Before proving our first result on almost everywhere summability we need a lemma.

LEMMA 9. *Let $N_p$ be a double Nörlund method of summability. Suppose there exists a constant $a>0$ such that*

$$(8.01)\qquad \sum_{j=1}^{n} j\,|\,p_j^{(k)} - p_{j-1}^{(k)}\,|\left(\frac{n}{j}\right)^a = O(|\,P_n^{(k)}\,|),\qquad k = 1, 2,$$

*and*

$$(8.02)\qquad \sum_{j=1}^{n} \frac{|\,P_j^{(k)}\,|}{j}\left(\frac{n}{j}\right)^a = O(|\,P_n^{(k)}\,|),\qquad k = 1, 2.$$

*Let $\lambda \geqq 1$ be any fixed number. Let*

$$(8.03)\qquad h_\lambda(x, y; f) = \sup_{m,n} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\,f(x+t, y+u)N_m^{(1)}(t)N_n^{(2)}(u)\,|\,dtdu$$

*where $m, n \geqq 1$, $m/n \leqq \lambda$, $n/m \leqq \lambda$. Then for any $\xi > 0$*

$$(8.04)\qquad \left|\,\underset{(x,y)}{E}\,\{[(x, y) \in Q][h_\lambda(x, y; f) > \xi]\}\,\right| \leqq \frac{A(a)\lambda^a}{\xi}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} |\,f(x, y)\,|\,dxdy$$

*where $A(a)$ depends only on $a$.*

**Proof.** If (8.01) and (8.02) hold for any $a = a_0 > 0$, then they also hold for all $a$ such that $0 \leqq a \leqq a_0$. Hence we may suppose that $0 < a < 1$.

Let $j$, $k$ be integers such that $2^j \leqq m < 2^{j+1}$, $2^k \leqq n < 2^{k+1}$. Also let $m/n \leqq \lambda$, $n/m \leqq \lambda$. Then $2^{|j-k|} \leqq 2\lambda$. Let

$$\begin{aligned} \overset{*}{t}_{mn}(x, y) &= \int_0^{\pi}\int_0^{\pi} |\,f(x+t, y+u)N_m^{(1)}(t)N_n^{(2)}(u)\,|\,dtdu \\ (8.05)\qquad &= \Big(\int_0^{\pi 2^{-j}}\int_{\pi 2^{-k}}^{\pi} + \int_{\pi 2^{-j}}^{\pi}\int_0^{\pi 2^{-k}} + \int_{\pi 2^{-j}}^{\pi}\int_{\pi 2^{-k}}^{\pi} \\ &\qquad + \int_0^{\pi 2^{-j}}\int_0^{\pi 2^{-k}}\Big) |\,f(x+t, y+u)N_m^{(1)}(t)N_n^{(2)}(u)\,|\,dudt \\ &= P_{jk}(x, y) + Q_{jk}(x, y) + R_{jk}(x, y) + S_{jk}(x, y). \end{aligned}$$

On account of Lemma 1, the estimates of §3 may be applied. Then from (3.03) and (3.12) we have

$$\begin{aligned} P_{jk}(x, y) \leqq \frac{Am}{R_n^{(2)}} \sum_{r=0}^{k-1} \int_0^{\pi 2^{-j}} dt \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} \frac{|\,f(x+t, y+u)\,|}{u} \\ \cdot \left\{R^{(2)}\left(\frac{1}{u}\right) + \frac{1}{u}\left[r_n^{(2)} + V_n^{(2)} - V^{(2)}\left(\frac{1}{u}\right)\right]\right\} du. \end{aligned}$$

By (7.01) and (7.04) the $r$th term of the sum on the right does not exceed

$$
\begin{aligned}
&\frac{Am}{R_n^{(2)}}\, 2^r\left\{R^{(2)}\left(\frac{2^{r+1}}{\pi}\right)+2^r\left[r_n^{(2)}+V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)\right]\right\}\\
&\qquad\cdot\int_{-\pi 2^{-j}}^{\pi 2^{-j}}dt\int_{-\pi 2^{-r}}^{\pi 2^{-r}}|f(x+t,y+u)|\,du\\
(8.06)\qquad &\qquad\leqq\frac{Am2^{-j}}{R_n^{(2)}}\left\{R_{2^r}^{(2)}\cdot 2^{a|j-r|}\right.\\
&\qquad\qquad\left.+\,2^{r+a|j-r|}\left[r_n^{(2)}+V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)\right]\right\}f^{*a}(x,y).
\end{aligned}
$$

In order to sum these terms we shall need the following:

$$
(8.07)\qquad \sum_{r=0}^{k-1}R_{2^r}^{(i)}\cdot 2^{a(k-r)}\leqq AR_{2^k}^{(i)},\qquad i=1,2,
$$

$$
(8.08)\qquad \sum_{r=0}^{k-1}2^{r+a(k-r)}\leqq A2^k,
$$

$$
(8.09)\qquad \sum_{r=0}^{k-1}2^{r+a(k-r)}\left[V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)\right]\leqq AR_n^{(2)}.
$$

To prove (8.07) we apply (3.06) to (8.02) and make use of Lemma 8. (8.08) is immediate. For (8.09) we first note that

$$
V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)=V_n^{(2)}=\sum_{s=1}^{n}|p_s^{(2)}-p_{s-1}^{(2)}|,\qquad r=0,1,
$$

$$
V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)\leqq V_n^{(2)}-V^{(2)}(2^{r-2})=\sum_{s=2^{r-2}+1}^{n}|p_s^{(2)}-p_{s-1}^{(2)}|,
$$

$$
r=2,3,4,\cdots,k-1.
$$

Substituting in the left side of (8.09), reversing the order of the summations and denoting the greatest integer less than or equal to $2+\log_2(s-1)$ by $g(s)$, we have

$$
\begin{aligned}
&\sum_{r=0}^{k-1}2^{r+a(k-r)}\left[V_n^{(2)}-V^{(2)}\left(\frac{2^r}{\pi}\right)\right]\\
&\qquad\leqq\sum_{s=2}^{n}|p_s^{(2)}-p_{s-1}^{(2)}|\sum_{r=2}^{g(s)}2^{r+a(k-r)}+\sum_{s=1}^{n}(2^{ak}+2^{1+a(k-1)})|p_s^{(2)}-p_{s-1}^{(2)}|\\
&\qquad=2^{ak}\sum_{s=2}^{n}|p_s^{(2)}-p_{s-1}^{(2)}|\,\frac{2^{(1-a)(g(s)+1)}-1}{2^{1-a}-1}+2^{ak}(1+2^{1-a})|p_1^{(2)}-p_0^{(2)}|\\
&\qquad\leqq An^a\sum_{s=1}^{n}s^{1-a}|p_s^{(2)}-p_{s-1}^{(2)}|.
\end{aligned}
$$

Then (8.09) follows from (8.01).

Summing (8.06) from $r=0$ to $k-1$, considering separately the cases $j\geqq k$ and $j<k$, and using (8.07)–(8.09) we get

(8.10) $$P_{jk}(x, y) \leqq A\lambda^a f^{*a}(x, y).$$

In the same way we obtain

(8.11) $$Q_{jk}(x, y) \leqq A\lambda^a f^{*a}(x, y).$$

Next from (3.12) we have

$$R_{jk}(x, y) \leqq A \sum_{r,s=0}^{j-1,k-1} Z_{rs}$$

where

$$Z_{rs} = \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} dt \int_{\pi 2^{-s-1}}^{\pi 2^{-s}} \left| f(x+t, y+u) \right| \left\{ M_{m1}^{(1)}(t) + M_{m2}^{(1)}(t) + M_{m3}^{(1)}(t) \right\}$$
$$\cdot \left\{ M_{n1}^{(2)}(u) + M_{n2}^{(2)}(u) + M_{n3}^{(2)}(u) \right\} du.$$

Each term of this sum consists of 9 parts each of which may be summed by making use of (8.07)–(8.09) and the analogue of (8.09). For example, let us consider that part arising from $M_{m1}^{(1)}(t)M_{n3}^{(2)}(u)$. The general term in this sum does not exceed

$$\frac{A}{R_m^{(1)}R_n^{(2)}} 2^{s+a|r-s|} R_{2^r}^{(1)} \left[ V_n^{(2)} - V^{(2)}\left(\frac{2^s}{\pi}\right) \right] f^{*a}(x, y).$$

Considering the case $j\geqq k$ we have

$$\sum_{r,s=0}^{j-1,k-1} 2^{s+a|r-s|} R_{2^r}^{(1)} \left[ V_n^{(2)} - V^{(2)}\left(\frac{2^s}{\pi}\right) \right]$$
$$\leqq \sum_{s=0}^{k-1} 2^s \left[ V_n^{(2)} - V^{(2)}\left(\frac{2^s}{\pi}\right) \right] \left\{ \sum_{r=0}^{s-1} R_{2^r}^{(1)} \cdot 2^{a(j-r)} + \sum_{r=s}^{j-1} R_{2^r}^{(1)} \cdot 2^{a(j-s)} \right\}$$
$$\leqq \sum_{s=0}^{k-1} 2^s \left[ V_n^{(2)} - V^{(2)}\left(\frac{2^s}{\pi}\right) \right] \sum_{r=0}^{j-1} R_{2^r}^{(1)} \cdot 2^{a(j-r)}$$
$$+ 2^{a(j-k)} \sum_{s=0}^{k-1} 2^{s+a(k-s)} \left[ V_n^{(2)} - V^{(2)}\left(\frac{2^s}{\pi}\right) \right] \sum_{r=0}^{j-1} R_{2^r}^{(1)}$$
$$\leqq A\lambda^a R_m^{(1)} R_n^{(2)}.$$

The case $k>j$ can be treated similarly. Thus it follows that

(8.12) $$R_{jk}(x, y) \leqq A\lambda^a f^{*a}(x, y).$$

Finally by (3.03)

$$S_{jk}(x, y) \leqq Amn \int_{-\pi 2^{-j}}^{\pi 2^{-j}} dt \int_{-\pi 2^{-k}}^{\pi 2^{-k}} |f(x+t, y+u)| \, du \tag{8.13}$$
$$\leqq Amn \cdot 2^{-j-k+a|j-k|} f^{*a}(x, y) \leqq A\lambda^a f^{*a}(x, y).$$

Combining (8.10)–(8.13) we see that $t^*_{mn}(x, y) \leqq A\lambda^a f^{*a}(x, y)$. But the integral on the right in (8.03) is the sum of four integrals, all analogous to $t^*_{mn}(x, y)$. Thus $h_\lambda(x, y; f) \leqq A\lambda^a f^{*a}(x, y)$. (8.04) now follows directly from Lemma 7.

9. **Restricted Nörlund summability almost everywhere.** We are now ready to prove our first theorem on almost everywhere summability.

THEOREM 5. *Let $N_p$ be a double Nörlund method of summability. Suppose there exists a constant $a > 0$ such that* (8.01) *and* (8.02) *are satisfied. Then $\sigma(f)$ is restrictly summable $N_p$ almost everywhere to $f$.*

**Proof.** This theorem follows immediately from Lemma 9. It suffices to make a decomposition $f = f_1 + f_2$ where $f_1$ is a trigonometrical polynomial and $f_2$ is such that

$$\left| \underset{(x,y)}{E} \{ |f_2(x, y)| > \delta \} \right| < \delta,$$

and

$$\left| \underset{(x,y)}{E} \{ \limsup |t_{mn}(x, y; f_2)| > \delta \} \right| < \delta$$

$(m/n \leqq \lambda,\ n/m \leqq \lambda,\ \lambda \geqq 1$ any fixed number), where $\delta$ is a fixed positive number as small as we please. Since $t_{mn}(x, y; f_1) \to f_1(x, y)$ it follows that $\limsup |t_{mn}(x, y; f) - f(x, y)|$ where $m, n \to \infty$ in such a manner that $m/n \leqq \lambda$, $n/m \leqq \lambda$ does not exceed $2\delta$ except on a set of measure less than $2\delta$. This completes the proof of the theorem.

The result of Marcinkiewicz and Zygmund [6], namely that $\sigma(f)$ is restrictedly summable $(C, \alpha, \beta)$, $\alpha, \beta > 0$, almost everywhere to $f$, follows immediately from Theorem 5.

10. **Lemma for square partial sums.** Turning now to the almost everywhere summability of the square partial sums we require the following lemma.

LEMMA 10. *Let $N_p$ be a regular Nörlund method of summability. Suppose there exists a constant $a > 0$ such that*

$$\sum_{j=1}^{n} j \,|p_j - p_{j-1}| \left(\frac{n}{j}\right)^a = O(|P_n|) \tag{10.01}$$

*and*

$$\sum_{j=1}^{n} \frac{|P_j|}{j} \left(\frac{n}{j}\right)^a = O(|P_n|). \tag{10.02}$$

*Let*

$$(10.03)\quad h^*(x, y; f) = \sup_{1\leqq n<\infty} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | f(x+t, y+u)K_n(t, u) | \, dtdu.$$

*Then for any* $\xi > 0$

$$(10.04)\quad \left| \underset{(x,y)}{E} \{[(x, y) \in Q][h^*(x, y; f) > \xi]\} \right| \leqq \frac{A(a)}{\xi} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} | f(x, y) | \, dxdy$$

*where* $A(a)$ *depends only on* $a$.

**Proof.** As in Lemma 9 we may suppose $0<a<1$. Let $k$ be an integer such that $2^k \leqq n < 2^{k+1}$. Let $D$ be the part of $Q$ $(-\pi, \pi; -\pi, \pi)$ in which $t, u \geqq 0$. Let $D^{(0)}$ be the part of $Q$ in which $t, u > \pi/2$. Divide $D - D^{(0)}$ into 9 domains $D_k^{(i)}$ $(i = 1, 2, 3, \cdots, 9)$ as in the proof of (3.16), the only difference being that in all the inequalities defining the regions $1/n$ is replaced by $\pi 2^{-k-1}$. We shall evaluate separately the integrals

$$(10.05)\quad \begin{aligned} A_k^{(i)} &= \iint_{D_k^{(i)}} | f(x+t, y+u)K_n(t, u) | \, dtdu, \quad i = 1, 2, 3, \cdots, 9, \\ A^{(0)} &= \iint_{D^{(0)}} | f(x+t, y+u)K_n(t, u) | \, dtdu. \end{aligned}$$

This may be done by methods similar to those used in the evaluation of $P_{jk}(x, y)$ and so on in Lemma 9. First of all from (3.01), (7.01) and (7.04) we get

$$(10.06)\quad A_k^{(1)} \leqq Af^{*a}(x, y).$$

In $D_k^{(2)}$ we note that $u \leqq t/2$, $t-u \geqq t/2 \geqq \pi 2^{-k-1} > 1/n$, $1/n \leqq t+u \leqq 2t$. Applying (3.11) and using the relations (8.07)–(8.09) we easily get

$$(10.07)\quad A_k^{(2)} \leqq Af^{*a}(x, y).$$

In $D_k^{(4)}$, $t \geqq u \geqq t/2$, $t+u \leqq 2t \leqq 4u$. Applying (3.02) and making the transformation $t = 2^{-1/2}(t'-u')$, $u = 2^{-1/2}(t'+u')$ we have

$$\begin{aligned} A_k^{(4)} \leqq A \sum_{r=0}^{k} \int_{-\pi 2^{-k-1}}^{0} du \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} \left| f\left(x + \frac{t-u}{2^{1/2}}, \ y + \frac{t+u}{2^{1/2}}\right) \right| t^{-2} dt \\ + A \int_{-\pi 2^{-k-1}}^{0} du \int_{\pi}^{\pi 2^{1/2}} \left| f\left(x + \frac{t-u}{2^{1/2}}, \ y + \frac{t+u}{2^{1/2}}\right) \right| t^{-2} dt. \end{aligned}$$

By (7.02) and (7.05) and taking account of (8.08) we have

$$(10.08)\quad A_k^{(4)} \leqq Af^{**a}(x, y).$$

In $D_k^{(6)}$, $u \leqq t/2$, $t-u \geqq t/2 \geqq \pi 2^{-k-1} > 1/n$, $1/n \leqq t+u \leqq 2t$. In $D_k^{(8)}$, $t-u \geqq 1/n$, $t \geqq u \geqq t/2$, $1/n \leqq t+u \leqq 2t \leqq 4u$. Then by (3.10) we have

$$A_k^{(6)} + A_k^{(8)} \leqq J_1 + J_2 + J_3$$

where

$$J_1 = \frac{A}{R_n}\iint_{D_k^{(6)}+D_k^{(8)}} \frac{|f(x+t,\ y+u)|}{tu}\left\{R\left(\frac{1}{t}\right) + \frac{1}{t}\left[r_n + V_n - V\left(\frac{1}{2t}\right)\right]\right\} dtdu,$$

$$J_2 = \frac{A}{R_n}\iint_{D_k^{(8)}} \frac{|f(x+t,\ y+u)|}{tu}\left\{R\left(\frac{2}{t}\right) + \frac{2}{t}\left[r_n + V_n - V\left(\frac{1}{t}\right)\right]\right\} dtdu,$$

$$J_3 = \frac{A}{R_n}\iint_{D_k^{(8)}} \frac{|f(x+t,\ y+u)|}{(t+u)^2}\left\{R\left(\frac{1}{t-u}\right) + \frac{1}{t-u}\left[r_n + V_n - V\left(\frac{1}{t-u}\right)\right]\right\} dtdu.$$

Now it is clear that

$$\iint_{D_k^{(6)}+D_k^{(8)}} (\cdots)dtdu \leqq \sum_{s=1}^{k}\sum_{r=0}^{s}\int_{\pi 2^{-s-1}}^{\pi 2^{-s}} du \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} (\cdots)dt,$$

$$\iint_{D_k^{(8)}} (\cdots)dtdu \leqq \sum_{s=1}^{k}\sum_{r=0}^{s-1}\int_{\pi 2^{-s-1}}^{\pi 2^{-s}} du \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} (\cdots)dt.$$

Using these facts and (8.07)–(8.09) we find that $J_1 + J_2 \leqq Af^{*a}(x, y)$. Transforming $J_3$ by the substitution $t = 2^{-1/2}(t'-u')$, $u = 2^{-1/2}(t'+u')$, noting that

$$\iint_{D_k^{(8)}} (\cdots)dt'du' \leqq \sum_{s=1}^{k+1}\sum_{r=0}^{s-1}\int_{-\pi 2^{-s}}^{-\pi 2^{-s-1}} du' \int_{\pi 2^{-r-1}}^{\pi 2^{-r}} (\cdots)dt'$$

$$+ \int_{-\pi/2}^{-\pi/4} du' \int_{\pi}^{\pi 2^{1/2}} (\cdots)dt',$$

and using (8.07)–(8.09) we find that $J_3 \leqq Af^{**a}(x, y)$. Thus

$$A_k^{(6)} + A_k^{(8)} \leqq A\{f^{*a}(x, y) + f^{**a}(x, y)\}. \tag{10.09}$$

Considering now the symmetric domains we have also

$$A_k^{(3)} + A_k^{(5)} + A_k^{(7)} + A_k^{(9)} \leqq A\{f^{*a}(x, y) + f^{**a}(x, y)\}. \tag{10.10}$$

In $D^{(0)}$, $t$, $u > \pi/2$. Applying (3.02), (7.01) and (7.04) we get

$$A^{(0)} \leqq Af^{*a}(x, y). \tag{10.11}$$

Combining (10.06)–(10.11) and noting that $Q$ is the sum of four domains like $D$ we have

$$h^*(x, y; f) \leqq A\{f^{*a}(x, y) + f^{**a}(x, y)\}. \tag{10.12}$$

(10.04) now follows immediately from Lemma 7.

**11. Summability of the square partial sums almost everywhere.**

THEOREM 6. *Let $N_p$ be a regular Nörlund method of summability. Suppose there exists a constant $a>0$ such that* (10.01) *and* (10.02) *are satisfied. Then the sequence $\{s_{nn}(x, y; f)\}$ is summable $N_p$ almost everywhere to $f(x, y)$.*

**Proof.** This theorem follows immediately from Lemma 10 just as Theorem 5 follows from Lemma 9.

We easily obtain the following corollary of Theorem 6.

COROLLARY 6.1. *The sequence $\{s_{nn}(x, y; f)\}$ is summable $(C, \alpha)$, $\alpha>0$, almost everywhere to $f(x, y)$.*

In conclusion let us note that conditions (5.02) and (6.02) do not in general hold almost everywhere. Hence it was not possible to deduce any results concerning almost everywhere summability from Theorems 1 and 3. However $(C, \alpha)$ applied to the square partial sums of the double Fourier series is effective almost everywhere if $\alpha>0$ but possesses the localization property if and only if $\alpha\geqq 1$. Also restricted $(C, \alpha, \beta)$ summability of the double Fourier series is effective almost everywhere if $\alpha, \beta>0$, but possesses the localization property if and only if $\alpha\geqq 1, \beta\geqq 1$.

## BIBLIOGRAPHY

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.

2. G. Grünwald, *Zur Summabilitätstheorie der Fourierschen Doppelreihe*, Proceedings of the Cambridge Philosophical Society, vol. 35 (1939), pp. 343–350.

3. ———, *Über die Summabilität der Fourierschen Reihe*, Acta Litterarum ac Scientiarum Szeged, Sectio Scientiarum Mathematicarum, vol. 10 (1941), pp. 55–63.

4. E. Hille and J. D. Tamarkin, *On the summability of Fourier series*. I, these Transactions, vol. 34 (1932), pp. 757–783.

5. J. Marcinkiewicz, *Sur une méthode remarquable de sommation des séries doubles de Fourier*, Annali della R. Scuola Normale Superiore di Pisa, (2), vol. 8 (1939), pp. 149–160.

6. J. Marcinkiewicz and A. Zygmund, *On the summability of double Fourier series*, Fundamenta Mathematicae, vol. 32 (1939), pp. 122–132.

BROWN UNIVERSITY,
PROVIDENCE, R. I.

# TRANSITIVITIES OF BETWEENNESS

BY

EVERETT PITCHER AND M. F. SMILEY

**Introduction.** The examination of the foundations of geometry which interested many prominent mathematicians about the turn of the century brought to light the importance of the fundamental notion of betweenness (see, for example[1], [10, 11]). This notion has suffered the treatment which modern mathematics metes out to all its concepts, namely, first an examination of the concept in a particular instance followed by wider and wider generalizations. The first part of this program for the concept of betweenness was carried through by Pasch, Huntington and Kline [8, 10]. The simplicity of the concept permitted them to give an elegant and complete theory for the case of linear order. In the direction of generalizations[2], K. Menger and his students have been one of the most important influences in the study of betweenness in metric spaces [9, 3].

We purpose here to add to both phases of this program. The first part of our paper continues the analysis of Huntington and Kline into an examination of postulates involving five points; the second part deals mainly with a definition of betweenness in lattices which generalizes metric betweenness in metric lattices (see [5, 6]). It is hoped that the five point transitivities may prove interesting and their analysis valuable. If we restrict our attention to the relation of betweenness in linear order such cannot be the case since four point properties are then sufficient to describe completely the betweenness relation. We feel that the results of the second part exhibit the properties of the betweenness relation as reflections of properties of the underlying space[3].

We shall use the notations of set theory which have become standard. In the second part we shall assume a knowledge of the fundamentals of both lattice theory and metric geometry. We refer the reader to the recent books *Distance Geometry* by L. M. Blumenthal [3] and *Lattice Theory* by Garrett Birkhoff [1]. We shall use the terminology and notation of these books in the second part.

---

Presented to the Society, May 2, 1941; received by the editors June 11, 1941.

(1) The numbers enclosed in brackets refer to the list of references at the end of the paper.

(2) The *chordal systems* recently introduced by W. Kaplan (Duke Mathematical Journal, vol. 7 (1940), pp. 165–167) are a generalization of linear order involving *two* triadic relations.

(3) The oft-quoted remark of K. Menger that Postulate B of Huntington and Kline should not be regarded as a property of betweenness but as a property of the underlying space [9, p. 79; 3, p. 36] indicates that it is easy to lose sight of this fact.

## Table of contents

Part I. Transitivities of betweenness



Part II. Applications



## Part I

We shall extend the discussion of an abstract relation of "betweenness" initiated by Pasch [10] and developed by Huntington and Kline [8] by relaxing some of the fundamental postulates of Huntington and Kline and by considering other possible postulates, particularly transitivities on five points.

1. **Fundamental assumptions.** We consider a set $K$ of points $a, b, c, d, x, \cdots$, and a triadic relation called *betweenness*, which holds (is *positive*) or fails (is *negative*) for each ordered triple of points, not necessarily distinct, in $K$. If the relation holds for the triple $a$, $b$, $c$, we write $abc$, read as written or as "$b$ is between $a$ and $c$." We make the following assumptions throughout Part I.

$\alpha$. *$abc$ if and only if $cba$* (symmetry in the end points).

$\beta$. *$abc$ and $acb$ if and only if $b=c$* (closure).

Postulate $\alpha$ is Postulate A of Huntington and Kline [8]. Postulate $\beta$ is similar to their Postulate C. Postulates $\alpha$ and $\beta$ together imply the statements (1) and (2) below.

(1) *$aba$ if and only if $a=b$.*

(2) *Every two positive relations on three points (not necessarily distinct) are equivalent or inconsistent.*

We do not assume that of an unordered triple of points one is between the other two (Postulate B of Huntington and Kline). In this respect our development will differ materially from theirs. This difference is essential because our interest lies in applications to lattices and metric spaces where B fails for very simple examples. We have replaced their Postulate D which requires that the three points of a linear triple be pairwise distinct by $\beta$ because we wish our five point transitivities to specialize under identification of two points to four point transitivities. This change, though logically necessary, is essentially only a change in terminology.

2. **Four point transitivities.** The statements about four points in which two positive relations of betweenness imply a third, which are theorems about linear order, and from which no hypothesis can be deleted leaving an equivalent statement, will be termed ***strong transitivities on four points.*** They are as follows.

$t_1$. $abc \cdot adb \rightarrow dbc$.

$t_2$. $abc \cdot adb \rightarrow adc$.

$t_3$. $abc \cdot bcd \cdot b \neq c \rightarrow abd$.

These are postulates (3), (2), and (1), respectively, of Huntington and Kline and are completely discussed by them. We shall need the fact that the only implication among the three is: "$t_1 \cdot t_3 \rightarrow t_2$" [8, p. 321].

3. **Weak transitivities on four points.** The statements concerning four points in which three distinct positive relations of betweenness imply a fourth, which are theorems about linear order, and from which no hypothesis may be deleted leaving an equivalent statement, will be termed ***weak transitivities on four points.*** They are as follows.

$\tau_1$. $abc \cdot adb \cdot adc \rightarrow dbc$.

$\tau_2$. $abc \cdot adb \cdot dbc \rightarrow adc$.

THEOREM 3.1. *The statements $\tau_1$ and $\tau_2$ are the only weak transitivities on four points. The implications $t_1 \rightarrow \tau_1$, $t_2 \rightarrow \tau_2$ hold.*

**Proof.** The second assertion is trivial. In order to prove the first assertion, we first observe that no two relations in hypothesis or conclusion of such a statement involve the same three letters by virtue of condition (2) of §1. Next, there are four ways of selecting unordered triples from four letters. Let $abc$ be the first hypothesis and $a, b, d$ be the letters in the second. Since $a, b, d$ or $b, c, d$ must occur in one hypothesis, it is always possible to achieve this situation by renaming the points. The second hypothesis is then one of the three, (1) $dab$, (2) $adb$, or (3) $abd$. The third hypothesis is on the points (i) $b, c, d$ or (ii) $a, c, d$. We examine the possible relations on the three letters in one of the sets (i), (ii) with (1), (2), or (3) for consistency with linear order; we then examine the other one of the sets (i), (ii) for a conclusion of a theorem about linear order. In the eight cases we have:

| | | *Third hypothesis* | *Conclusion* |
|---|---|---|---|
| (1) | (i) | $bcd$ | $acd$ |
| (1) | (ii) | $acd$ | $bcd$ |
| (2) | (i) | $dbc$ | $adc$ |
| (2) | (ii) | $adc$ | $dbc$ |
| (3) | (i) | $bcd$ | $acd$ |
| (3) | (i) | $bdc$ | $adc$ |
| (3) | (ii) | $acd$ | $bcd$ |
| (3) | (ii) | $adc$ | $bdc$ |

It is readily seen that these eight theorems reduce to two on suitable permutations of the letters of the letters of the hypotheses and conclusions. These two are $\tau_1$ and $\tau_2$. This completes the proof.

From this discussion of weak transitivities on four points it is apparent that an attempt at a weaker statement about four points with four or more hypotheses and one conclusion must contain two hypotheses or a hypothesis and a conclusion identical under $\alpha$ or a hypothesis or conclusion true under $\beta$.

4. **Five point transitivities.** The statements concerning five points in which three positive relations of betweenness imply a fourth, which are theorems about linear order, and from which no hypothesis can be deleted leaving an equivalent statement, will be termed *strong transitivities on five points*. They are as follows.

$T_1$. $abc \cdot adb \cdot xdb \cdot b \neq d \rightarrow xdc.$
$T_2$. $abc \cdot adb \cdot bcx \cdot b \neq c \rightarrow dcx.$
$T_3$. $abc \cdot adb \cdot xcd \cdot c \neq d \rightarrow acx.$
$T_4$. $abc \cdot dab \cdot xcd \rightarrow abx.$
$T_5$. $abc \cdot adc \cdot bxd \rightarrow axc$
$T_6$. $abc \cdot adb \cdot acx \rightarrow dcx.$
$T_7$. $abc \cdot abd \cdot cxd \rightarrow abx.$
$T_8$. $abc \cdot dab \cdot xcd \cdot a \neq b \rightarrow acx.$
$T_9$. $abc \cdot dab \cdot xcd \cdot a \neq b \rightarrow bcx.$
$T_{10}$. $abc \cdot abd \cdot xbc \cdot a \neq b \cdot b \neq c \rightarrow xbd.$

* * *

$T_{11}$. $abc \cdot adc \cdot xab \cdot a \neq b \rightarrow xad.$
$T_{12}$. $abc \cdot adc \cdot xab \cdot a \neq b \rightarrow xdc.$
$T_{13}$. $abc \cdot adb \cdot xab \cdot a \neq b \rightarrow xdc.$
$T_{14}$. $abc \cdot adb \cdot xac \rightarrow xad.$
$T_{15}$. $abc \cdot adb \cdot xac \rightarrow xdb.$
$T_{16}$. $abc \cdot adb \cdot xac \rightarrow xdc.$
$T_{17}$. $abc \cdot adb \cdot acx \rightarrow adx.$
$T_{18}$. $abc \cdot adb \cdot acx \rightarrow dbx.$
$T_{19}$. $abc \cdot adb \cdot xad \cdot a \neq d \rightarrow xac.$
$T_{20}$. $abc \cdot adb \cdot xad \cdot a \neq d \rightarrow xbc.$
$T_{21}$. $abc \cdot adb \cdot xad \cdot a \neq d \rightarrow xdc.$
$T_{22}$. $abc \cdot adb \cdot bxc \rightarrow adx.$
$T_{23}$. $abc \cdot adb \cdot bxc \rightarrow dbx.$
$T_{24}$. $abc \cdot adb \cdot bcx \cdot b \neq c \rightarrow adx.$
$T_{25}$. $abc \cdot adb \cdot bcx \cdot b \neq c \rightarrow dbx.$
$T_{26}$. $abc \cdot adb \cdot bdx \cdot b \neq d \rightarrow xbc.$
$T_{27}$. $abc \cdot adb \cdot xcd \cdot c \neq d \rightarrow abx.$
$T_{28}$. $abc \cdot adb \cdot xcd \cdot c \neq d \rightarrow adx.$
$T_{29}$. $abc \cdot adb \cdot xcd \rightarrow dbx.$

$T_{30}$. $abc \cdot adb \cdot xcd \rightarrow bcx$.
$T_{31}$. $abc \cdot adb \cdot cxd \rightarrow axc$.
$T_{32}$. $abc \cdot adb \cdot cxd \rightarrow adx$.
$T_{33}$. $abc \cdot adb \cdot cdx \rightarrow xbc$.
$T_{34}$. $abc \cdot adb \cdot cdx \rightarrow xdb$.
$T_{35}$. $abc \cdot dab \cdot adx \cdot a \neq b \cdot a \neq d \rightarrow xac$.
$T_{36}$. $abc \cdot dab \cdot adx \cdot a \neq b \cdot a \neq d \rightarrow xbc$.
$T_{37}$. $abc \cdot dab \cdot xcd \cdot a \neq b \rightarrow dax$.
$T_{38}$. $abc \cdot dab \cdot xcd \cdot a \neq b \rightarrow dbx$.

THEOREM 4.1. *The statements $T_1$–$T_{38}$ are a complete list of strong transitivities on five points.*

**Proof.** In order to effect this enumeration we reason as follows. One letter in the three hypotheses must occur only once since there are nine places to be filled with five letters. We shall denote this letter (or one such if there are more) by $x$ and agree that it occurs in the third hypothesis. Then the first two hypotheses are on four letters with two letters in common. Letting $abc$ be the first hypothesis and $d$ the remaining letter, we see that the second hypothesis must be on the letters (i) $a, c, d$ or (ii) $a, b, d$; the case $b, c, d$ reduces to $a, b, d$ on interchange of $a$ and $c$, which by virtue of $\alpha$ does not change $abc$. Calling $a$ and $c$ in $abc$ *terminal* and $b$ *medial*, we see that the letters common to the first and second hypotheses must fall under one of the following cases:

I. each letter terminal in both hypotheses;

II. one letter terminal in both; one letter medial once (say in the first) and terminal once;

III. one letter terminal in both; one medial in both;

IV. each letter terminal once and medial once.

Possible pairs of hypotheses to fill the first two places are then

A. *abc and adc*, C. *abc and abd*,
B. *abc and adb*, D. *abc and dab*.

On examination one will find that we have used the pairs I (i), II (ii), III (ii), and IV (ii). The pairs I (ii), II (i), III (i), and IV (i) are incompatible with linear order. With any of the pairs A, B, C, D we can use a third hypothesis on

$$(4.1)\qquad \begin{array}{llllll} (1) & a, b, x, & (2) & a, c, x, & (3) & a, d, x, \\ (4) & b, c, x, & (5) & b, d, x, & (6) & c, d, x. \end{array}$$

Every letter which occurs only once in the hypotheses must occur in the conclusion. For, if we had a theorem about linear order for which this were false,

we would obtain an equivalent one by dropping the hypothesis containing the letter; and we have agreed not to consider statements with this property.

In the six subcases (1)–(6) under A we can use a conclusion on

$$(4.2)\qquad \begin{array}{l} (1) \text{ or } (4)\ a, d, x;\ b, d, x;\ c, d, x. \\ (2)\ b, d, x. \\ (3)\ a, b, x;\ b, c, x;\ b, d, x. \\ (5)\ a, b, x;\ a, c, x;\ a, d, x;\ b, c, x;\ c, d, x. \\ (6)\ a, b, x;\ b, c, x;\ b, d, x. \end{array}$$

In the six subcases (1)–(6) under B, C, or D we can use a conclusion on

$$(4.3)\qquad \begin{array}{l} (1)\ c, d, x. \\ (2) \text{ or } (4)\ a, d, x;\ b, d, x;\ c, d, x. \\ (3) \text{ or } (5)\ a, c, x;\ b, c, x;\ c, d, x. \\ (6)\ a, b, x;\ a, c, x;\ a, d, x;\ b, c, x;\ b, d, x. \end{array}$$

We proceed then to an examination of cases according to the following plan. With each pair A, B, C, D we inspect the three arrangements of the letters in the six cases of (4.1) to see whether they are consistent with linear order. With each consistent arrangement we inspect each of the three arrangements of letters in the corresponding set of (4.2) and (4.3) to determine whether a theorem in linear order is obtained. The work is shortened by examining each set of three hypotheses as we proceed to see whether it has already occurred under some permutation of the letters; this examination is facilitated by applying the classification scheme I–IV to the three pairs of hypotheses.

This procedure yields the transitivities $T_1$–$T_{38}$ though not in that order, and the proof is complete.

The reader will observe that the program initiated in the determination of the transitivities $\tau_1$, $\tau_2$, and $T_1$–$T_{38}$ could be extended to include transitivities with more hypotheses and more letters. We shall not do this.

5. **Selection of fundamental five point transitivities.** Each of the transitivities $T_{11}$–$T_{38}$ is equivalent to a combination of the transitivities $t_1$, $t_2$, $t_3$. We shall state and prove these facts in the following form

$T_{11}.\ \sim.\ t_1 \cdot t_3\ [b=c;\ c=d.\ abc,\ xab\ (t_3)\ xac,\ adc\ (t_1)\ xad.]$

We mean that $T_{11}$ is equivalent to $t_1$ and $t_3$; that $t_1$ is proved by identifying $b$ and $c$; that $t_3$ is proved by identifying $c$ and $d$; that $T_{11}$ is proved from $t_1$ and $t_3$ by applying $t_3$ to $abc$ and $xab$ to obtain $xac$, and by applying $t_1$ to $xac$ and $adc$ to obtain $xad$. We use $\alpha$ and $\beta$ freely without explicit reference. Whenever we use $t_3$ the letters common to the two hypotheses are distinct as required.

$T_{12}$. $\sim.\ t_2 \cdot t_3$ [$b=c$; $a=d$. $abc$, $xab$ ($t_3$) $xac$, $adc$ ($t_2$) $xdc$.]
$T_{13}$. $\sim.\ t_2 \cdot t_3$ [$b=c$; $a=d$. $abc$, $adb$ ($t_2$) $adc$; $abc$, $xab$ ($t_3$) $xac$, $adc$ ($t_2$) $xdc$.]
$T_{14}$. $\sim.\ t_1$ [$b=c$. $abc$, $xac$ ($t_1$) $xab$, $adb$ ($t_1$) $xad$.]
$T_{15}$. $\sim.\ t_1 \cdot t_2$ [$a=d$; $b=c$. $abc$, $xac$ ($t_1$) $xab$, $adb$ ($t_2$) $xdb$.]
$T_{16}$. $\sim.\ t_2$ [$a=x$. $abc$, $adb$ ($t_2$) $adc$, $xac$ ($t_2$) $xdc$.]
$T_{17}$. $\sim.\ t_2$ [$c=x$. $abc$, $adb$ ($t_2$) $adc$, $acx$ ($t_2$) $adx$.]
$T_{18}$. $\sim.\ t_1 \cdot t_2$ [$b=c$; $a=d$. $abc$, $acx$ ($t_2$) $abx$, $adb$ ($t_1$) $dbx$.]
$T_{19}$. $\sim.\ t_3$ [$b=c$. $adb$, $xad$ ($t_3$) $xab$, $abc$ ($t_3$) $xac$.]
$T_{20}$. $\sim.\ t_3$ [$b=d$. $adb$, $xad$ ($t_3$) $xab$, $abc$ ($t_3$) $xbc$.]
$T_{21}$. $\sim.\ t_2 \cdot t_3$ [$a=x$; $b=c$. $abc$, $adb$ ($t_2$) $adc$, $xad$ ($t_3$) $xdc$.]
$T_{22}$. $\sim.\ t_1 \cdot t_2$ [$b=d$; $c=x$. $abc$, $bxc$ ($t_1$) $abx$, $adb$ ($t_2$) $adx$.]
$T_{23}$. $\sim.\ t_1$ [$a=d$. $abc$, $adb$ ($t_1$) $dbc$, $bxc$ ($t_1$) $dbx$.]
$T_{24}$. $\sim.\ t_2 \cdot t_3$ [$c=x$; $b=d$. $abc$, $bcx$ ($t_3$) $abx$, $adb$ ($t_2$) $adx$.]
$T_{25}$. $\sim.\ t_1 \cdot t_3$ [$c=x$; $a=d$. $abc$, $bcx$ ($t_3$) $abx$, $adb$ ($t_1$) $dbx$.]
$T_{26}$. $\sim.\ t_1 \cdot t_3$ [$d=x$; $a=d$. $abc$, $adb$ ($t_1$) $dbc$, $bdx$ ($t_3$) $xbc$.]
$T_{27}$. $\sim.\ t_2 \cdot t_3$ [$a=d$; $b=c$. $abc$, $adb$ ($t_2$) $adc$, $dcx$ ($t_3$) $acx$, $abc$ ($t_2$) $abx$.]
$T_{28}$. $\sim.\ t_2 \cdot t_3$ [$c=x$; $b=c$. $abc$, $adb$ ($t_2$) $adc$, $xcd$ ($t_3$) $adx$.]
$T_{29}$. $\sim.\ t_1 \cdot t_2$ [$c=x$; $a=d$. $abc$, $adb$ ($t_1$) $dbc$, $xcd$ ($t_2$) $dbx$.]
$T_{30}$. $\sim.\ t_1$ [$a=d$. $abc$, $adb$ ($t_1$) $dbc$, $xcd$ ($t_1$) $bcx$.]
$T_{31}$. $\sim.\ t_2$ [$b=c$. $abc$, $adb$ ($t_2$) $adc$, $cxd$ ($t_2$) $axc$.]
$T_{32}$. $\sim.\ t_1 \cdot t_2$ [$b=c$; $c=x$. $abc$, $adb$ ($t_2$) $adc$, $cxd$ ($t_1$) $adx$.]
$T_{33}$. $\sim.\ t_1 \cdot t_2$ [$d=x$; $a=d$. $abc$, $adb$ ($t_1$) $dbc$, $cdx$ ($t_2$) $xbc$.]
$T_{34}$. $\sim.\ t_1$ [$a=d$. $abc$, $adb$ ($t_1$) $dbc$, $cdx$ ($t_1$) $xdb$.]
$T_{35}$. $\sim.\ t_3$ [$b=c$. $abc$, $dab$ ($t_3$) $cad$, $adx$ ($t_3$) $xac$.]
$T_{36}$. $\sim.\ t_3$ [$d=x$. $dab$, $adx$ ($t_3$) $xab$, $abc$ ($t_3$) $xbc$.]
$T_{37}$. $\sim.\ t_2 \cdot t_3$ [$b=c$; $c=x$. $abc$, $dab$ ($t_3$) $dac$, $xcd$ ($t_2$) $dax$.]
$T_{38}$. $\sim.\ t_2 \cdot t_3$ [$a=d$; $c=x$. $abc$, $dab$ ($t_3$) $dbc$, $xcd$ ($t_2$) $dbx$.]

We summarize the results of this section in the following theorem.

THEOREM 5. *Each of the transitivities $T_{11}$–$T_{38}$ is equivalent to a combination of the transitivities $t_1$, $t_2$, and $t_3$.*

**6. The logical relations among the fundamental strong transitivities.** None of the transitivities $T_1$–$T_{10}$ is equivalent to a combination of the transitivities $t_1$, $t_2$, $t_3$. We shall devote this section and the following one to a proof of this fact. In addition we shall construct the essentials of a complete existential theory of $t_1$–$t_3$, $T_1$–$T_{10}$. The basic implications are given in this section. We use the notation explained in §5.

$t_1 \cdot t_3 \rightarrow T_1 \rightarrow t_2 \cdot t_3$ [$abc$, $adb$ ($t_1$) $dbc$, $xdb$ ($t_3$) $xdc$. $a=x$; $a=d$.]
$t_1 \cdot t_3 \rightarrow T_2 \rightarrow t_3$ [$abc$, $adb$ ($t_2$) $adc$; $abc$, $bcx$ ($t_3$) $acx$, $adc$ ($t_1$) $dcx$. $a=d$.]
$t_2 \cdot t_3 \rightarrow T_3 \rightarrow t_3$ [$abc$, $adb$ ($t_2$) $adc$, $xcd$ ($t_3$) $acx$. $b=c$.]

$t_1 \cdot t_3 \to T_4 \to t_1 \cdot t_2$ [*abc*, *dab* ($t_3$) *dbc*, *xcd* ($t_2$) *dbx*, *dab* ($t_1$) *abx*. $b=c$; $a=d$.]
$T_5 \to t_2$ [$a=b$.]
$t_1 \cdot t_2 \to T_6 \to t_1$ [*abc*, *adb* ($t_2$) *adc*, *acx* ($t_1$) *dcx*. $b=c$.]
$T_7 \to t_1$ [$b=d$.]
$t_1 \cdot t_3 \to T_8 \to t_1$ [*abc*, *dab* ($t_3$) *dac*, *xcd* ($t_1$) *acx*. $b=c$.]
$t_1 \cdot t_3 \to T_9 \to t_1$ [*abc*, *dab* ($t_3$) *dbc*, *xcd* ($t_1$) *bcx*. $a=d$.]

It is apparent that when two letters of a statement $T_1$–$T_{10}$ are identified, the resulting statement is either a tautology or is equivalent to one of the statements $t_1$, $t_2$, $t_3$, $\tau_1$ or $\tau_2$. We may see that we cannot thus obtain either $\tau_1$ or $\tau_2$ as follows. Notice that the hypotheses of both $\tau_1$ and $\tau_2$ contain one letter three times and three letters twice and that the conclusion of each is on these latter three letters. Suppose that an identification of two letters leads to $\tau_1$ or $\tau_2$. Then $x$ must be identified with some letter because it occurs only once in the hypotheses of each of $T_1$–$T_{10}$. Since $x$ always appears in the conclusions of $T_1$–$T_{10}$, the letter to be identified with $x$ can occur only once in the hypotheses. It must then also occur in the conclusion along with $x$. By virtue of $\beta$ the conclusion is then either vacuous or implies still further identification. This contradicts our assumption that one of the statements $\tau_1$ or $\tau_2$ appears on identifying two letters.

*The above list of implications includes all nontautological results obtained by identifying two letters*. This fact will be useful in simplifying the examination of the table of examples to be given in §7.

In the proofs of the following implications, the results of the preceding implications are used.

$T_3 \to T_6$ [*abc*, *adb* ($t_1$) *dbc*, *adb*, *acx*, $d \neq b$ ($T_3$) *dcx*.
If $d=b$ then *abc*, *acx* ($t_1$) *dcx*.]
$T_3 \to T_9$ [*abc*, *dab*, *xcd*, $a \neq b$ ($T_3$) *acx*, *abc* ($t_1$) *bcx*.]
$T_4 \cdot T_9 \to T_3$ [*abc*, *dab*, *xcd* ($T_4$) *abx*; *abc*, *dab*, *xcd*, $a \neq b$ ($T_9$) *bcx*; *abx*, *bcx* ($t_2$) *acx*.]
$T_7 \cdot T_9 \to T_3$ [*abc*, *dab*, *xcd*, $a \neq b$ ($T_9$) *bcx*, *dab*, *xcd* ($T_7$) *xca*.]
$t_2 \cdot T_3 \to T_4$ [*abc*, *dab*, *xcd*, $a \neq b$ ($T_3$) *acx*, *abc* ($t_2$) *abx*. If $a=b$, $T_4$ is true.]
$t_1 \cdot T_7 \to T_6$ [*abc*, *adb* ($t_1$) *dbc*; *abc*, *acx* ($t_1$) *bcx*, *acx*, *adb* ($T_7$) *dcx*.]
$t_2 \cdot t_3 \cdot T_{10} \to T_1$ [*abc*, *adb* ($t_2$) *adc*, *adb*, *xdb*, $a \neq d$, $b \neq d$ ($T_{10}$) *xdc*.
If $a=d$, *abc*, *xab* ($t_3$) *xdc*.]

We shall devote the next section to the proof of the following theorem.

THEOREM 6. *The implications listed in this section are the only ones holding among the statements $t_1$–$t_3$, $T_1$–$T_{10}$.*

REMARK. It seems to be worth mentioning that $t_1 \cdot t_2 \cdot t_3 \to T_1$, $T_2$, $T_3$, $T_4$, $T_6$, $T_8$, $T_9$; but that $t_1 \cdot t_2 \cdot t_3$ *does not imply* $T_5$, $T_7$, or $T_{10}$ (for proof see §7). We are of the opinion that the interest of a five point transitivity varies inversely

as its logical intimacy with $t_1$, $t_2$, and $t_3$. Viewed in this light, $T_{10}$ is surely the most interesting—but we still lack a "concrete" interpretation for it.

7. **The examples in the existential theory.** We shall complete the existential theory begun in §6. We have attempted to make our list of examples as simple as possible through the use of composite examples. No attempt has been made to make the number of points in each example the least possible [14, p. 250].

The following elementary examples will be used in the table which concludes this section. In each of the examples the positive relations are those listed together with the ones which follow from $\alpha$ and $\beta$. In the first four examples the class $K$ consists of four distinct points, while in the remaining examples $K$ consists of five distinct points. Certain of these examples are merely the statement of the hypotheses of one of the transitivities. We indicate this by giving the example the same number as the transitivity, replacing $t$ by $k$, $\tau$ by $\kappa$ and $T$ by $K$.

$k1$. *abc adb.*
$k3$. *abc dab.*
$\kappa 1$. *abc adb adc.*
$\kappa 2$. *abc adb dbc.*
$K3$. *abc adb xcd.*
$K4$. *abc dab xcd.*
$K5$. *abc adc bxd.*
$K7$. *abc abd cxd.*
$K10$. *abc abd xbc.*
$E1$. *abc dab xcd abx.*
$E2$. *abc dab xcd bcx.*
$E3$. *abc dab xcd acx bcx.*
$E4$. *abc adb acx dbc bcx.*
$E5$. *abc adb bcx adc acx abx adx.*
$E6$. *abc adb xdb adc.*
$E7$. *abc adb bcx abx acx.*

The following table of examples completes our existential theory. In entry 4 we take as the space $K$ the points $a, b, c, d, x, a', b', c', d', x'$ with the positive relations of example $K5$ on the points $a, b, c, d, x$, the positive relations of example $K7$ on the points $a', b', c', d', x'$, and the other positive relations required by $\beta$. Each case in which the example column contains more than one entry is to be treated similarly. We have made no column for $T_{10}$. It will be found to hold in each of our examples except 32–35, where it must fail because of the implication $t_2 \cdot t_3 \cdot T_{10} \rightarrow T_1$. To secure the example corresponding to those listed in which $T_{10}$ fails we simply adjoin $K10$ to the example listed.

| | $t_1$ | $t_2$ | $t_3$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | $T_7$ | $T_8$ | $T_9$ | *Example* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1. | + | + | + | + | + | + | + | + | + | + | + | + | Linear order |
| 2. | | | | | | | | + | | − | | | $K7$ |
| 3. | | | | | | | | − | | + | | | $K5$ |
| 4. | | | | | | | | − | | − | | | $K5$, $K7$ |
| 5. | + | + | − | − | − | − | + | + | + | + | + | + | $k3$ |
| 6. | | | | | | | + | + | | + | − | − | $E1$ |
| 7. | | | | | | | + | + | | − | + | + | $k3$, $K7$ |
| 8. | | | | | | | + | + | | − | − | − | $E1$, $K7$ |
| 9. | | | | | | | − | + | | + | − | − | $K4$ |
| 10. | | | | | | | − | + | | − | − | + | $E2$ |
| 11. | | | | | | | − | + | | − | − | − | $K4$, $K7$ |
| 12–18. | | | | | | | * | − | | * | * | * | as in 5–11 with $K5$ |
| 19. | + | − | − | − | − | − | − | − | + | + | + | + | $E3$ |
| 20. | | | | | | | | | + | + | − | − | $K4$, $\kappa2$ |
| 21. | | | | | | | | | + | − | + | + | $E3$, $K7$ |
| 22. | | | | | | | | | + | − | − | + | $E2$, $\kappa2$ |
| 23. | | | | | | | | | + | − | − | − | $K4$, $K7$, $\kappa2$ |
| 24. | | | | | | | | | − | − | − | + | $E4$ |
| 25. | | | | | | | | | − | − | − | − | $E4$, $K4$ |
| 26. | − | + | + | + | + | + | − | + | − | − | − | − | $\kappa1$ |
| 27. | | | | + | + | | | − | | | | | $\kappa1$, $K5$ |
| 28. | | | | + | − | | | + | | | | | $E5$ |
| 29. | | | | + | − | | | − | | | | | $E5$, $K5$ |
| 30. | | | | − | + | | | + | | | | | $E6$ |
| 31. | | | | − | + | | | − | | | | | $E6$, $K5$ |
| 32. | | | | − | − | | | + | | | | | $E5$, $E6$ |
| 33. | | | | − | − | | | − | | | | | $E5$, $E6$, $K5$ |
| 34. | − | + | − | − | − | − | − | + | − | − | − | − | $\kappa1$, $k3$ |
| 35. | | | | | | | | − | | | | | $\kappa1$, $k3$, $K5$ |
| 36. | − | − | + | − | + | + | − | − | − | − | − | − | $k1$ |
| 37. | | | | | + | − | | | | | | | $K3$ |
| 38. | | | | | − | + | | | | | | | $E7$ |
| 39. | | | | | − | − | | | | | | | $E7$, $K3$ |
| 40. | − | − | − | − | − | − | − | − | − | − | − | − | $k1$, $k3$ |

* In these places arrange + and − signs as in the entries 5–11.

## PART II

We shall devote the remainder of this paper to the study of a generalization of metric betweenness in metric lattices, and to the application of the transitivities of Part I both to this relation and to the relation of betweenness in semi metric, metric, and metric ptolemaic spaces.

8. **Lattice betweenness.** Glivenko [5, 6] proved that in a metric lattice an element $b$ is (metrically) between the elements $a$ and $c$ if and only if

$$(8.1) \qquad (a \cap b) \cup (b \cap c) = b = (a \cup b) \cap (b \cup c).$$

This condition does not involve the metric and we take it as our definition(4) of *betweenness* in an arbitrary lattice $L$. When $b$ is between $a$ and $c$ we shall frequently write simply $abc$. We shall need the simple and fundamental properties of this relation given in the following two lemmas(5).

LEMMA 8.1. *If $L$ is a lattice and $a, b, c \in L$, then*
(1) *the inequalities $a \leqq b \leqq c$ imply that the relation $abc$ holds;*
(2) *the relation $abc$ implies that $a \cap c \leqq b \leqq a \cup c$;*
(3) *both $a \cap c$ and $a \cup c$ are between $a$ and $c$.*

**Proof.** (1) If $a \leqq b \leqq c$, then $(a\cap b)\cup(b\cap c)=a\cup b=b=b\cap c=(a\cup b)\cap(b\cup c)$. By our definition, $abc$; and (1) is proved.

(2) If $abc$, then $b\cap(a\cap c)=(a\cup b)\cap(b\cup c)\cap(a\cap c)=a\cap c$. It follows that $a\cap c\leqq b$. Dually, $b\leqq a\cup c$. This proves (2).

(3) Note that $(a\cap(a\cup c))\cup((a\cup c)\cap c)=a\cup c$, and also that $(a\cup(a\cup c))\cap((a\cup c)\cup c)=a\cup c$ By definition, $a\cup c$ is between $a$ and $c$. Dually, $a\cap c$ is between $a$ and $c$. This proves (3).

LEMMA 8.2. *If $L$ is a lattice then its betweenness relation satisfies $\alpha$ and $\beta$.*

**Proof.** ($\alpha$) This is an immediate consequence of the commutativity of the operations $a\cap b$ and $a\cup b$ in lattices.

($\beta$) Let $L$ be a lattice containing elements $a, b, c$ for which the relations $abc$ and $acb$ hold. We then have $b=(a\cup b)\cap(b\cup c)$ and $c=(a\cup c)\cap(c\cup b)$, and hence $b\cap c=(a\cup b)\cap(b\cup c)\cap(c\cup a)$. Consequently,

$$b \cap c \geqq (a \cup b) \cap (b \cup c) \cap b = b \geqq b \cap c,$$
$$b \cap c \geqq c \cap (b \cup c) \cap (c \cup a) = c \geqq b \cap c.$$

It follows that $b\cap c=b=c$. To prove the converse we must show that $aac$ is valid in lattices for every pair of elements $a, c$. It is easily seen that $(a\cap a)\cup(a\cap c)=a\cup(a\cap c)=a$. Using duality we see by the definition that the relation $aac$ holds. The proof is complete.

In addition to these fundamental properties, we now show that lattice betweenness possesses the five point transitivity $T_6$.

THEOREM 8.1. *If $L$ is a lattice then its betweenness relation satisfies the transitivity $T_6$.*

(4) G. Birkhoff [1, p. 9] also gives a definition of *betweenness* which applies to partially ordered sets and which has all the transitivities of Part I.

(5) We may also note that $abc$ holds if and only if $a\cap c\leqq b\leqq a\cup c$ and $(a, b, c)D$ (see J. von Neuman, *Continuous Geometries*, Princeton Lecture Notes, 1936–1937; and Lemma 10.1 below).

**Proof.** Let $L$ be a lattice and consider elements $a, b, c, d, x \in L$ for which the relations $abc$, $adb$, and $acx$ hold. We wish to show that $dcx$ is true. We prove first that $(d \cap c) \cup (c \cap x) = c$. Notice that $c = (a \cap c) \cup (c \cap x)$; and, by Lemma 8.1 (2), that $a \cap c \leqq b$, and $a \cap b \leqq d$. It follows that $a \cap c \leqq a \cap b \leqq d$. We obtain

$$\begin{aligned}(d \cap c) \cup (c \cap x) &= (d \cap ((a \cap c) \cup (c \cap x))) \cup (c \cap x) \\ &\geqq (d \cap a \cap c) \cup (d \cap c \cap x) \cup (c \cap x) \\ &\geqq (a \cap c) \cup (d \cap c \cap x) \cup (c \cap x) \\ &\geqq (a \cap c) \cup (c \cap x) \\ &\geqq c \geqq (d \cap c) \cup (c \cap x).\end{aligned}$$

Consequently, $(d \cap c) \cup (c \cap x) = c$. Dually, $(d \cup c) \cap (c \cup x) = c$. By definition, we have $dcx$. The proof is complete.

COROLLARY. *The transitivities $t_1$ and $\tau_1$ are valid for the betweenness relation in every lattice.*

**Proof.** This is a trivial result of the implications $T_6 \rightarrow t_1 \rightarrow \tau_1$.

9. **Interpretations of certain of the five point transitivities for lattice betweenness.** Glivenko [5] showed that a metric lattice is distributive if and only if its (metric) betweenness relation has the transitivity which we have labeled $T_6$. We shall extend this result to lattice betweenness in this section. We shall also prove that both $T_4$ and $T_7 \cdot t_2$ are equivalent to the distributive law; that each of the transitivities $t_2$ and $\tau_2$ is equivalent to the modular law; and that each one of the postulates $t_3$, $T_1$, $T_2$, and $T_3$ holds if and only if the lattice is linearly ordered. The remaining transitivities do not seem to have important lattice-theoretic interpretations. We shall verify that each of them fails in the Boolean algebra of eight elements.

Our first theorem gives the interpretation of the transitivity $t_2$.

THEOREM 9.1. *A lattice $L$ is modular if and only if its betweenness relation satisfies the transitivity $t_2$.*

**Proof.** Consider first a modular lattice $L$ containing elements $a, b, c, d$ for which the relations $abc$ and $adb$ hold. We wish to establish the relation $adc$. Note that $(a \cap b) \cup (b \cap c) = b$, $(a \cap d) \cup (d \cap b) = d$, and, by Lemma 8.1 (2), that $a \cap c \leqq b$, and $a \cap b \leqq d$. We then obtain $d = (a \cap d) \cup (d \cap b) = (a \cap d) \cup (d \cap ((a \cap b) \cup (b \cap c)))$. Using the modular law, since $a \cap b \leqq d$, we find that

$$\begin{aligned} d &= (a \cap d) \cup (a \cap b) \cup (d \cap b \cap c) \\ &= (a \cap d) \cup (d \cap b \cap c) \\ &\leqq (a \cap d) \cup (d \cap c) \leqq d. \end{aligned}$$

Hence $d = (a \cap d) \cup (d \cap c)$. Dually, $d = (a \cup d) \cap (d \cup c)$. Consequently, the relation $adc$ is valid. Thus the modular law implies the transitivity $t_2$. Conversely,

the transitivity $t_2$ implies the modular law. To see this, let $L$ be a lattice whose betweenness satisfies $t_2$. If $L$ is non-modular it must contain the simplest non-modular lattice of five elements shown in Figure 9.1 as a sublattice.

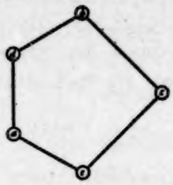

FIG. 9.1

Note that, by Lemma 8.1 (3), the relation $abc$ holds in $L$ since $b=a\cup c$; and that the relation $adb$ holds in $L$ by Lemma 8.1 (1). But if the relation $adc$ is true, then $(a\cap d)\cup(d\cap c)=d$. However, we see from Figure 9.1 that $(a\cap d)\cup(d\cap c)=a\neq d$. Thus the transitivity $t_2$ fails in $L$. This is contrary to our hypothesis. It follows that the lattice $L$ is modular. The proof is complete.

A similar result holds for the transitivity $\tau_2$.

THEOREM 9.2. *If $L$ is a lattice, then its betweenness relation satisfies the transitivity $\tau_2$ if and only if $L$ is modular.*

**Proof.** If $L$ is a modular lattice, it is clear from the implication $t_2\rightarrow\tau_2$ and Theorem 9.1 that $\tau_2$ is valid for the betweenness of $L$. On the other hand, if $\tau_2$ holds then the lattice must be modular. Otherwise a sublattice such as we have pictured in Figure 9.1 exists. In it we have shown, in the proof of Theorem 9.1, that the relations $abc$ and $adb$ are true and that the relation $adc$ is false. But the relation $dbc$ also holds in the lattice of Figure 9.1, since $b=d\cup c$. Hence the hypotheses of the transivity $\tau_2$ hold in this sublattice (and therefore also in the lattice itself), but its conclusion fails. This is contrary to the assumption that the transivity $\tau_2$ holds. Thus the transitivity $\tau_2$ implies that the modular law is valid. The proof is complete.

We pass now to a discussion of the transitivity $T_5$. Our next lemma, on the road to establishing the equivalence of $T_5$ and the distributive law, gives a relation between Duthie's *segments*(6) and our *betweenness*.

LEMMA 9.1. *If $L$ is a lattice then it is distributive if and only if for every triple $a, b, c\in L$ the inequalities $a\cap c\leqq b\leqq a\cup c$ imply that the relation $abc$ holds.*

**Proof.** Consider a lattice $L$ in which the implication of our lemma holds. We establish the modular law for $L$ first. Consider three elements $a, b, c\in L$ with $a\leqq c$. Since $a\cap b\leqq c\cap(a\cup b)\leqq a\cup b$, our hypothesis yields that $c\cap(a\cup b)$ is between $a$ and $b$. Whence we have

(6) Duthie defines a *segment* of a lattice $L$ between two elements $a, b\in L$ as the set of all $x\in L$ satisfying $a\cap b\leqq x\leqq a\cup b$. Our lemma has also been proved by him [4].

$$\begin{aligned} c \cap (a \cup b) &= (a \cap c \cap (a \cup b)) \cup (c \cap (a \cup b) \cap b) \\ &= (a \cap c) \cup (c \cap b) = a \cup (b \cap c), \end{aligned}$$

which is the modular law. Now consider elements $u$, $v$, $w \in L$. Note that $(u \cap w) \leqq (u \cap w) \cup (v \cap (u \cup w)) \leqq (u \cup w)$. By hypothesis we obtain that $z \equiv (u \cap w) \cup (v \cap (u \cup w))$ is between $u$ and $w$. An easy application of the modular law reduces the conditions that this relation hold to the equations

$$(u \cap v) \cup (v \cap w) \cup (w \cap u) = z = (u \cup v) \cap (v \cup w) \cap (w \cup u).$$

But this last identity characterizes distributive lattices [1, p. 74]. Conversely, if $L$ is distributive and $a$, $b$, $c$ are three elements of $L$ for which $a \cap c \leqq b \leqq a \cup c$, then $(a \cap b) \cup (b \cap c) = b \cap (a \cup c) = b$, and dually. Thus the relation $abc$ holds. This completes the proof.

We continue with the proof that $T_5$ is equivalent to the distributive law.

THEOREM 9.3. *If $L$ is a lattice, then its betweenness relation has the transitivity $T_5$ if and only if $L$ is distributive.*

**Proof.** Consider a lattice $L$ whose betweenness relation satisfies $T_5$. By Lemma 9.1, $L$ will be distributive provided the relation $abc$ holds for every triple $a$, $b$, $c \in L$ such that $a \cap c \leqq b \leqq a \cup c$. Hence consider elements $a$, $b$, $c \in L$ for which $a \cap c \leqq b \leqq a \cup c$. By Lemma 8.1 (2), $b$ is between $a \cap c$ and $a \cup c$. By Lemma 8.1 (3), we know that both $a \cap c$ and $a \cup c$ are between $a$ and $c$. Application of the transitivity $T_5$ then yields the fact that $b$ is between $a$ and $c$. Thus the validity of the transitivity $T_5$ implies the distributive law by Lemma 9.1. Conversely, if $L$ is distributive and the relations $abc$, $adc$, and $bxd$ hold for elements $a$, $b$, $c$, $d$, $x \in L$, then, using Lemma 8.1 (2), we obtain

$$a \cap c \leqq b \cap d \leqq x \leqq b \cup d \leqq a \cup c.$$

Since $L$ is distributive, it then follows from Lemma 9.1 that $b$ is between $a$ and $c$. Hence the distributive law implies that the transitivity $T_5$ holds in $L$. This completes the proof.

Still another form of the distributive law is provided by the postulate $T_4$, while $T_7$ is equivalent to the distributive law in *modular* lattices. The next three theorems will show this.

THEOREM 9.4. *If $L$ is a distributive lattice, then its betweenness relation has the transitivities $T_4$ and $T_7$.*

**Proof.** Let $L$ be a distributive lattice. We prove first that $T_4$ holds for the betweenness of $L$. Consider five elements $a$, $b$, $c$, $d$, $x \in L$ for which the relations $abc$, $dab$, and $xcd$ hold. We wish to prove that the relation $abx$ is valid. By Lemma 9.1 it is sufficient to show that $a \cap x \leqq b \leqq a \cup x$. Lemma 8.1 (2) yields that $a \leqq b \cup d$, and that $x \cap d \leqq c$. Hence we find that

$$a \cap x \leqq x \cap (b \cup d) = (x \cap b) \cup (x \cap d) \leqq (x \cap b) \cup c.$$

Combining with $a$ we have

$$a \cap x \leqq a \cap ((x \cap b) \cup c) = (a \cap x \cap b) \cup (a \cap c) \leqq (a \cap b) \cup (b \cap c).$$

But $(a\cap b)\cup(b\cap c)=b$, since the relation $abc$ holds. It follows that $a\cap x\leqq b$. Dually, $a\cup x\geqq b$. By Lemma 9.1 and the fact that $L$ is distributive we then know that the relation $abx$ is valid. Thus the transitivity $T_4$ is valid in distributive lattices.

To prove that $T_7$ holds for the betweenness relation of a distributive lattice $L$, consider five elements $a, b, c, d, x\in L$ for which the relations $abc$, $abd$, and $cxd$ hold. We wish to show that we have the relation $abx$. By Lemma 8.1 (2), we know that $a\cap c\leqq b$, $a\cap d\leqq b$, and that $x\leqq c\cup d$. Combining the last inequality with $a$, we find that $a\cap x\leqq a\cap(c\cup d)=(a\cap c)\cup(a\cap d)\leqq b$. Dually, $a\cup x\geqq b$. It follows from Lemma 9.1 that the relation $abx$ is true. Thus $T_7$ is valid in distributive lattices. The proof is complete.

THEOREM 9.5. *If $L$ is a lattice whose betweenness relation has the transitivity $T_4$, then $L$ is distributive.*

**Proof.** The implication $T_4\rightarrow t_2$, proved in §6, together with the result of Theorem 9.1 shows that if $T_4$ holds for lattice betweenness in a lattice $L$ then $L$ is modular. It is well known [1, p. 75] that every modular non-dis-

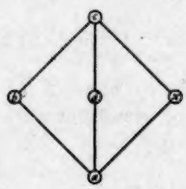

FIG. 9.2

tributive lattice contains a copy of the simplest modular non-distributive lattice shown in Figure 9.2 as a sublattice. Thus if $T_4$ were to hold for lattice betweenness in a non-distributive lattice $L$ it would hold in the lattice of Figure 9.2. In this lattice we have the relations $abc$, $dab$, and $xcd$ since $a<b<c$, $a=b\cap d$, and $c=d\cup x$. But $abx$ would require that $(a\cap b)\cup(b\cap x)=b$, while actually $(a\cap b)\cup(b\cap x)=a\cup a=a\neq b$. Thus $T_4$ fails in this lattice. It follows that the transitivity $T_4$ for lattice betweenness implies that the lattice is distributive. The proof is complete.

THEOREM 9.6. *If $L$ is a modular lattice whose betweenness relation satisfies the transitivity $T_7$ then $L$ is distributive.*

**Proof.** Since $L$ is modular it can fail to be distributive only if it has a sublattice of the type shown in Figure 9.2. If we reletter the elements of this

lattice, putting $a' = b$, $b' = a$, $c' = x$, $d' = d$, and $x' = c$, we may verify easily that the relations $a'b'c'$, $a'b'd'$, and $c'x'd'$ hold since $b' = a' \cap c'$, $b' = a' \cap d'$, and $x' = c' \cup d'$. If $T_7$ held we should have $(a' \cup b') \cap (b' \cup x') = b'$, while in fact

$$(a' \cup b') \cap (b' \cup x') = a' \cap x' = a' \neq b'.$$

Thus $T_7$ fails for $L$. It follows that a modular lattice $L$ cannot fail to be distributive when $T_7$ holds. The proof is complete.

REMARK. An examination of the lattice of Figure 9.1 will show that the result of Theorem 9.6 cannot be extended to non-modular lattices.

Our next theorem discusses the transitivities $T_1$, $T_2$, $T_3$, and $t_3$.

THEOREM 9.7. *If $L$ is a lattice then its betweenness relation has one of the transitivities $T_1$, $T_2$, $T_3$, $t_3$ if and only if $L$ is linearly ordered.*

**Proof.** Since the transitivities cited obviously hold in a linear order and since each of them implies $t_3$, it will suffice to show that the betweenness relation of a lattice satisfies $t_3$ only if the lattice is a linear order. Hence let $L$ be a lattice whose betweenness relation satisfies $t_3$. Consider two elements $a, b \in L$. Suppose that none of the relations $a < b$, $a > b$, $a = b$ holds. Then clearly $a \cup b \neq a$ and $a \cap b \neq a$. Note that $a \cap b < a < a \cup b$. By Lemma 8.1 (1), we find that $a$ is between $a \cup b$ and $a \cap b$. By Lemma 8.1 (3), $a \cup b$ is between $a$ and $b$. The transitivity $t_3$ then yields the fact that $a$ is between $a \cap b$ and $b$. It follows that

$$a = (b \cap a) \cup (a \cap (a \cap b)) = (a \cap b) \cup (a \cap b) = a \cap b,$$

contrary to the fact that $a \cap b \neq a$. Thus, if $t_3$ holds for the betweenness of $L$ no pair of elements of $L$ can be incomparable. This means that $L$ is linearly ordered. The proof is complete.

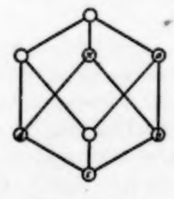

FIG. 9.3

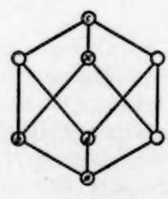

FIG. 9.4

We now show that each of the remaining transitivities, namely, $T_8$, $T_9$, and $T_{10}$, fail to hold in the Boolean algebra of eight elements. To see that $T_{10}$ fails note that in Figure 9.3 we have $abc$, $abd$, $xbc$, and $bxd$ since $a < b < c$, $a \cap d < b < a \cup d$, $c < b < x$, and $x = b \cup d$; but if $xbd$ also held then $\beta$ would require that $x = b$. Figure 9.4 provides a counterexample for $T_8$ and $T_9$. Using Lemma 8.1, we see that we have $abc$, $dab$, and $xcd$ since $a < b < c$, $a = d \cap b$,

and $c=d\cup x$. But $bcx$ is false since $b\cup x=x$, which does not contain $c$; and $acx$ is false since $a\cup x=x$.

Let us summarize the results of this section and the preceding one in a theorem.

THEOREM 9.8. *If $L$ is a lattice then its betweenness relation has the transitivities $T_6$, $t_1$, and $\tau_1$; it has each of the transitivities $t_2$ and $\tau_2$ if and only if $L$ is modular; it has each of the transitivities $T_4$ and $T_5$ if and only if $L$ is distributive; it has the transitivity $T_7$ if and only if $L$ is distributive provided that $L$ is modular; and it has each of the transitivities $t_3$, $T_1$, $T_2$, and $T_3$ if and only if $L$ is linearly ordered.*

10. **Critique of lattice betweenness.** A. Wald found a set of properties of metric betweenness which characterize this relation in metric spaces [13]. We shall devote this section to a proof of an analogous result for lattice betweenness. The algebraic structure of lattices permits a slight economy in that we may characterize our relation of lattice betweenness in the particular lattice considered, while Wald found it necessary to consider a relation $R$ defined in every metric space.

The present form of this section is due in large measure to suggestions of W. R. Transue. He, together with one of us, applies the result in a study of transitivities of betweenness in metric lattices and their generalizations.

Our result takes the following form.

THEOREM 10.1. *If $L$ is a lattice and $R$ is a triadic relation defined for all ordered*(7) *triples of elements of $L$, then $R$ is lattice betweenness provided that the following conditions hold.*

(i) *$R$ satisfies the postulates ($\alpha$) and ($\beta$).*

(ii) *$R$ satisfies the transitivity $t_1$.*

(iii) *If $a\leqq b\leqq c$, then $(a, b, c)R$.*

(iv) *The relations $(a, a\cup c, c)R$, $(a, a\cap c, c)R$ hold for every $a, c\in L$.*

(v) *If the relation $abc$ holds, then in the sublattice generated by $a$, $b$, $c$ the transitivity $t_2$ holds for $R$.*

The properties (i)–(iv) have already been established for lattice betweenness in Lemmas 8.1 and 8.2 and in the corollary to Theorem 8.1. The following lemma justifies the assumption (v).

LEMMA 10.1. *If $L$ is a lattice and the relation $abc$ holds for three elements $a, b, c\in L$, then the sublattice generated by $a$, $b$, $c$ is distributive.*

**Proof.** We shall prove, in fact, that the free lattice [1, p. 22] generated by three elements $a$, $b$, $c$ for which the relation $abc$ is assumed is given in

(7) This word "ordered" refers, of course, to the fundamental metamathematical notion of "ordered" set. This should not be confused with the order relation in the lattice $L$.

Figure 10.1. To prove this, let $a$, $b$, $c$ be three elements which generate a lat-

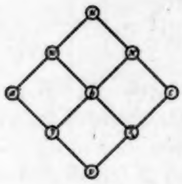

FIG. 10.1

tice in which the relation $abc$ holds. We note first that we must then have $b=(a\cup b)\cap(b\cup c)$. Hence we must also have

$$\begin{aligned}((a \cup b) \cap c) \cup b) &= ((a \cup b) \cap c) \cup ((a \cup b) \cap (b \cup c)) \\ &= (a \cup b) \cap (b \cup c) = b.\end{aligned}$$

It follows that $b \geqq (a\cup b)\cap c$. Consequently, $b\cap c \geqq (a\cup b)\cap c \geqq b\cap c$, and hence $b\cap c=(a\cup b)\cap c$. Using this result we see that $(b\cap c)\cup(a\cap c)= (c\cap(a\cup b))\cup(a\cap c)=c\cap(a\cup b)=b\cap c$. Interchange of $a$ and $c$ in these results and their duals justifies Figure 10.1. It is obvious that the lattice of Figure 10.1 is distributive since it is the product [1, p. 13 and p. 76] of two chains of three elements. The proof is complete.

**Proof of Theorem** 10.1. Consider a lattice $L$ and a triadic relation $R$ defined for all ordered triples of elements of $L$ which satisfies the conditions (i)–(v) of Theorem 10.1. We prove first that the relation $(a, b, c)R$ implies the relation $abc$. For this implication we need only the conditions (i)–(iv). Consider three elements $a, b, c\in L$ for which the relation $(a, b, c)R$ holds. By (iv) we have $(a, a\cup b, b)R$ and (ii) then gives $(a\cup b, b, c)R$. Again (iv) gives $(b, b\cup c, c)R$ and (ii) yields $(a\cup b, b, b\cup c)R$. Note that $b\leqq(a\cup b)\cap(b\cup c) \leqq a\cup b$, and apply (iii) to obtain $(b, (a\cup b)\cap(b\cup c), a\cup b)R$. Combining this last relation with $(a\cup b, b, b\cup c)R$ and using (ii) we find that $(b\cup c, b, (a\cup b)\cap(b\cup c))R$. But since $b\leqq(a\cup b)\cap(b\cup c)\leqq b\cup c$, (iii) gives $(b, (a\cup b)\cap(b\cup c), b\cup c)R$. Using (i) we then obtain $b=(a\cup b)\cap(b\cup c)$. By duality, $b=(a\cap b)\cup(b\cap c)$, and we conclude that the relation $abc$ holds. Thus the relation $(a, b, c)R$ implies the relation $abc$.

We prove next that the relation $abc$ implies the relation $(a, b, c)R$. For this implication we do not use the condition (ii). In the proof we shall omit explicit reference to our use of the condition (i). Let $a$, $b$, $c$ be three elements of $L$ for which the relation $abc$ holds. By Lemma 8.1 (2) we have $a\leqq a\cup b\leqq a\cup b\cup c =a\cup c$, and the relations $(a, a\cup b\cup c, c)R$ and $(a, a\cup b, a\cup b\cup c)R$ then follow from (iv) and (iii). Condition (v) then gives $(a, a\cup b, c)R$. Note that $c\leqq b\cup c\leqq a\cup b\cup c$. The relations $(c, b\cup c, a\cup b\cup c)R$ and $(c, a\cup b\cup c, a\cup b)R$ then follow from (iii) and (iv). Applying (v) we obtain $(c, b\cup c, a\cup b)R$. Since $abc$ holds we have $b=(a\cup b)\cap(b\cup c)$, and (iv) then gives $(b\cup c, b, a\cup b)R$.

Condition (v) then yields $(c, b, a\cup b)R$. Combining this last relation with $(c, a\cup b, a)R$ and using (v) again we find $(a, b, c)R$. Thus the relation $abc$ implies the relation $(a, b, c)R$.

Combining the results of the preceding two paragraphs we find that the relation $R$ holds if and only if lattice betweenness holds, that is, $R$ *is* the lattice betweenness of $L$. The proof of Theorem 10.1 is complete.

REMARK. It seems unfortunate that our theorem requires the condition (v). That it is necessary to make some such assumption may be seen by considering the lattice of Figure 10.1. In this lattice let $R$ be the same as lattice betweenness except that the relation $(a, b, c)R$ does not hold. If we could prove $(a, b, c)R$ from the assumptions (i)–(iv) of Theorem 10.1, then we should have to obtain this result from condition (ii) since the conclusions of (i), (ii), and (iii) cannot apply to a triple $(d, e, f)$ with both $d$ and $e$ and $f$ and $e$ not comparable. To obtain $(a, b, c)R$ from (ii) would require hypotheses of the form $(d, b, c)R$, $(d, a, b)R$ or of the form $(d, b, a)R$, $(d, c, b)R$. But these sets cannot hold in our example, since if we have $(d, a, b)R$ and $(d, b, c)R$, then $d\neq a$, and $d\cap b\leqq a\leqq d\cup b$. It follows that $d\cup b=w$ or $u$ and hence that $d\cap b=b$, contrary to $d\cap b\leqq a$. The other set of hypotheses may be treated likewise by interchanging $a$ and $c$. It is possible to give alternatives for the condition (v) but we shall not consider them here.

11. **Betweenness in metric, semi metric, and metric ptolemaic spaces.** In a metric space with distance function $\delta$ one says [3, p. 38] that $q$ is *between* the points $p$ and $r$ in case $\delta(p, q)+\delta(q, r)=\delta(p, r)$ and $p\neq q\neq r$. It is evident that this relation fails to satisfy our condition $\beta$. We suggest that it should be modified so as to satisfy $\beta$ by deleting the condition $p\neq q\neq r$ which requires that the points $p$, $q$, $r$ be pairwise distinct. We shall do this and shall write $pqr$ for the modified relation, reserving the locution "$q$ is between $p$ and $r$" for the usual relation. K. Menger [9] established the transitivities $t_1$ and $t_2$ for metric betweenness. His famous example of a "railroad" space [9, p. 80] was constructed to prove that the transitivity $T_5$ may fail in metric spaces. For the case of a semi metric space [3, p. 38] O. Taussky found that the weak transitivity $\tau_1$ holds for the analogue of metric betweenness. Examples of semi metric spaces are easily given in which $\tau_2$ fails.

There has recently been some interest in spaces which are metric and ptolemaic [12, 2], that is, metric spaces in which the three products of the lengths of opposite sides of every quadrilateral are the sides of some triangle in the euclidean plane. For such spaces L. M. Blumenthal [2] established the transitivity $t_3$. Thus in metric ptolemaic spaces we have immediately the properties $T_1$–$T_4$, $T_8$, and $T_9$. It is interesting that $T_5$ also holds in such spaces. We may see this as follows(8). Let $a$, $b$, $c$, $d$, $x$ be five points of a metric ptolemaic space which satisfy the relations $abc$, $adc$, and $bxd$. Using the ptole-

(8) Professor Blumenthal has also noted this fact in a letter to one of us. We are indebted to him for a stimulating correspondence during the preparation of this paper.

maic inequality(9) we obtain $ax\cdot bd \leqq ab\cdot xd + ad\cdot xb$ and $cx\cdot bd \leqq bc\cdot xd + dc\cdot xb$. Adding these inequalities we find

$$(11.1) \qquad bd(ax+cx) \leqq xd(ab+bc)+xb(ad+dc) = xd\cdot ac + xb\cdot ac = bd\cdot ac.$$

If $b=d$, then by (1) of §1, $b=d=x$, and the relation $axc$ is implied by the relation $abc$. If $b\neq d$, then $ax+cx=ac$ from (11.1) and the triangle inequality, and the relation $axc$ is true. As an example of the use of the relation $abc$ instead of "$b$ is between $a$ and $c$," let us give the proof of $T_8$ for the second relation. It will suffice to prove that $a\neq x\neq c$. Suppose that $a=x$. By hypothesis, $a$ is then between $b$ and $d$. The transitivity $t_3$ then gives ($a\neq b$ !) that $a$ is between $d$ and $c$, which contradicts $d$ between $a$ and $c$. If $x=c$, then by hypothesis $c$ is between $a$ and $d$. The transitivity $t_3$ then gives ($b\neq c$!) that $c$ is between $a$ and $d$, which contradicts $d$ between $a$ and $c$.

None of the remaining five points transitivities, namely, $T_6$, $T_7$, and $T_{10}$, holds in every metric ptolemaic space. This may be shown by examples of spaces consisting of five points.

## References


1. Garrett Birkhoff, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940.
2. L. M. Blumenthal, *Betweenness in metric ptolemaic spaces*, Bulletin of the American Mathematical Society, abstract 47-1-66.
3. ———, *Distance Geometries*, The University of Missouri Studies, vol. 13, no. 2 (1938).
4. W. D. Duthie, *Segments of ordered sets*, these Transactions, vol. 51 (1942), pp. 1–14.
5. V. Glivenko, *Contributions à l'étude des systèmes de choses normées*, American Journal of Mathematics, vol. 59 (1937), pp. 941–956.
6. ———, *Géométrie des systèmes de choses normées*, American Journal of Mathematics, vol. 58 (1936), pp. 799–828.
7. E. V. Huntington, *A new set of postulates for betweenness with proof of complete independence*, these Transactions, vol. 26 (1924), pp. 257–282.
8. E. V. Huntington and J. R. Kline, *Sets of independent postulates for betweenness*, these Transactions, vol. 18 (1917), pp. 301–325.
9. K. Menger, *Untersuchungen über allgemeine Metrik*, Mathematische Annalen, vol. 100 (1928), pp. 75–163.
10. Pasch-Dehn, *Grundlagen der Geometrie*, Berlin, 1926.
11. Gilbert de B. Robinson, *The Foundations of Geometry*, Mathematical Expositions, Toronto, no. 1, 1940.
12. I. J. Schoenberg, *Metric arcs of vanishing Menger curvature*, Annals of Mathematics, (2), vol. 41 (1940), pp. 715–726.
13. A. Wald, *Axiomatik des Zwischenbegriffes in metrischen Räume*, Mathematischen Annalen, vol. 104, (1930–1931), pp. 476–484.
14. W. E. van de Walle, *On the complete independence of the postulates for betweenness*, these Transactions, vol. 24 (1926), pp. 249–256.


Lehigh University,
Bethlehem, Pa.

(9) For convenience we now write the distance between two points $a$ and $b$ of the metric space simply as $ab$.

# ON THE BASIS THEOREM FOR DIFFERENTIAL SYSTEMS

BY
E. R. KOLCHIN

One of the principal points of departure in the study of polynomials and polynomial ideals is the Hilbert basis theorem, which states that every set $\mathfrak{m}$ of polynomials in a finite number of indeterminates contains a finite subset $f_1, \cdots, f_s$ such that

$$\mathfrak{m} \subseteq (f_1 \cdots f_s).$$

As originally proved by Hilbert, this theorem applied to polynomials whose coefficients were either elements of a field, or rational integers. In keeping with the modern tendency toward abstraction, however, one now finds the theorem proved for polynomials whose coefficients are elements of a commutative ring with unit element in which every set has a finite basis.

When one turns to differential polynomials and differential ideals one finds that the exact analogue of the Hilbert theorem is lacking(1). It is not true that every system of differential polynomials $\Sigma$ contains a finite subset $F_1, \cdots, F_s$ such that

$$\Sigma \subseteq [F_1 \cdots F_s](^2).$$

Instead one is forced to choose as a starting point a weakened analogue, the basis theorem of Ritt and Raudenbush. This theorem has been proved for differential polynomials in a finite number of unknowns (indeterminates) $y_1, \cdots, y_n$ with any differential field of characteristic zero as coefficient domain (3), and may be stated in either of the two following equivalent forms:

1. Every system $\Sigma$ of differential polynomials has a finite subset $F_1, \cdots, F_s$ such that, for each differential polynomial $A \in \Sigma$ there is a positive integer $t$ such that $A^t \in [F_1, \cdots, F_s]$.

2. Every system $\Sigma$ of differential polynomials has a finite subset $F_1, \cdots, F_s$ such that $\Sigma$ is contained in the perfect differential ideal generated by $F_1, \cdots, F_s$:

$$\Sigma \subseteq \{F_1 \cdots, F_s\}(^4).$$

---

Presented to the Society, February 22, 1941; received by the editors July 3, 1941.

(1) See J. F. Ritt, *Differential Equations from the Algebraic Standpoint*, American Mathematical Society Colloquium Publications, vol. 14, New York, 1932, pp. 12–13.

(2) Square brackets [ ] are used for differential ideals. Parentheses ( ) denote, as usual, (algebraic) ideals.

(3) See H. W. Raudenbush, these Transactions, vol. 36 (1934), pp. 361–368.

(4) The perfect differential ideal generated by a set is denoted by the set enclosed in braces { }.

That these two statements are equivalent (when the coefficient domain is a differential field of characteristic zero) follows from the fact that the set of all differential polynomials some powers of which are in $[F_1, \cdots, F_s]$ is a perfect differential ideal(5).

It is the object of the present paper to generalize the basis theorem of Ritt and Raudenbush, as the Hilbert basis theorem has been generalized, to permit more general coefficient domains. There is nothing in the literature, for example, which allows treatment of differential polynomials with the set of rational integers, or a differential field of nonvanishing characteristic, as the domain of coefficients.

An easy counterexample shows at the outset that there is no hope of generalizing the first statement of the theorem. In $\mathfrak{J}\{y\}$, the set of all ordinary differential polynomials in $y$ with rational integral coefficients, the system

$$y^p, y_1^p, y_2^p, \cdots$$

where $p$ is any integer greater than 1, is such a counterexample(6). For, no matter what $n$ is, no power of $y_{n+1}^p$ is contained in $[y^p, y_1^p, \cdots, y_n^p]$. This is easy to see since $y_{n+1}$ appears in $[y^p, y_1^p, \cdots, y_n^p]$ only in terms divisible by $p$ or by some $y_i^p$ $(i \leqq n)$.

On the other hand the second statement of the theorem above is susceptible of generalization, although not so wide a one as might be expected at first blush. A finite subset $b_1, \cdots, b_s$ of a subset $\phi$ of a differential ring $\mathfrak{R}$ is called a basis of $\phi$ if

$$\phi \subseteq \{b_1, \cdots, b_s\}.$$

If every subset of $\mathfrak{R}$ has a basis we say that the basis theorem holds in $\mathfrak{R}$. Our main theorem asserts that:

*If $\mathfrak{R}$ is a commutative differential ring with unit element, in which the basis theorem holds, and if $\mathfrak{R}$ also satisfies a certain condition termed "regularity," then the basis theorem holds in any commutative differential ring $\mathfrak{R}'$ obtained from $\mathfrak{R}$ by a finite number of differential ring adjunctions.* An example shows that the regularity condition is not superfluous.

The admittance of more general coefficient domains complicates the structure of perfect differential ideals and makes it desirable to represent, after Raudenbush, the perfect differential ideal $\{\phi\}$ generated by a set $\phi$ as the set-theoretic limit of a non-decreasing sequence of sets denoted by $\{\phi\}_n$. (See §1.) This permits the classification of some bases as 0-bases, 1-bases, 2-bases, and so on.

This naturally raises the question whether a set which has a basis has an

(5) Raudenbush, loc. cit., p. 363. Raudenbush neglects to state that the differential rings he considers must contain the rational number system.

(6) $y_j$ denotes the $j$th derivative of $y$.

$m$-basis for some $m$. This question is only partially answered below and still remains for investigation. If every set in a differential ring has an $m$-basis for some $m$ dependent on the set then we say that the $*$-basis theorem holds in that ring. What we show is that *if the $*$-basis theorem holds in $\mathfrak{R}$ then the $*$-basis theorem holds in $\mathfrak{R}'$* ($\mathfrak{R}$, $\mathfrak{R}'$ as above). Thus we see that every set of differential polynomials in $\mathfrak{J}\{y_1, \cdots, y_n\}$ has an $m$-basis for some $m$. However, it is still unknown whether we may put a bound on $m$. An example shows that any such bound would depend on $n$.

For the sake of generality the proofs are given for partial differential rings. There is a proof for ordinary differential rings which is materially shorter and simpler, and which is not a specialization of the partial case. For its own interest we present in §11 an outline of this proof.

1. **Perfect differential ideals.** Throughout this paper $\mathfrak{R}$ will denote a commutative (partial) differential ring with $r$ types of differentiation (or derivative operators) $\delta_1, \cdots, \delta_r$.

A differential ideal $\sigma$ in $\mathfrak{R}$ is called *perfect* if $\sigma$ contains an element of $\mathfrak{R}$ whenever it contains some power of that element: $a^t \in \sigma$ implies $a \in \sigma$.

Let $\phi$ be an arbitrary subset of $\mathfrak{R}$. There exists a perfect differential ideal in $\mathfrak{R}$ containing $\phi$; for example, $\mathfrak{R}$ itself. The intersection of all perfect differential ideals containing $\phi$ is itself a perfect differential ideal containing $\phi$, and is called the perfect differential ideal generated by $\phi$; in symbols, $\{\phi\}$.

To exhibit the structure of $\{\phi\}$ we define by induction:

$\{\phi\}_0 = (\phi)$,

$\{\phi\}_n =$ set of all $a \in \mathfrak{R}$ such that $a^t \in [\{\phi\}_{n-1}]$ for some $t$, $n = 1, 2, \cdots$.

Each $\{\phi\}_n$ is an ideal. When $n > 0$, $\{\phi\}_n$ contains every element some power of which it contains. Moreover,

$$\{\{\phi\}_m\}_n = \{\phi\}_{m+n}$$

and

$$\phi \subseteq (\phi) = \{\phi\}_0 \subseteq \{\phi\}_1 \subseteq \{\phi\}_2 \subseteq \cdots \subseteq \{\phi\}.$$

The definitions imply that

$$\{\phi\} = \{\phi\}_0 + \{\phi\}_1 + \{\phi\}_2 + \cdots \text{ (7)}.$$

2. **Bases.** A finite subset $b_1, \cdots, b_s$ of $\phi \subseteq \mathfrak{R}$ is called a *basis* of $\phi$ if

$$\phi \subseteq \{b_1, \cdots, b_s\}.$$

The basis will be called an *$m$-basis* if

(7) If $\mathfrak{R}$ is a differential ring obtained by the differential ring adjunction of a finite number of unknowns to a differential field of characteristic 0 then $\{\phi\} = \{\phi\}_1$, as is well known. For general $\mathfrak{R}$ this is no longer true. For example, if $\mathfrak{R}$ is the totality of differential polynomials in $y$ with rational integral coefficients, we see, because $y \in \{y^2\}$, that $y_1$, the derivative of $y$, is in $\{y^2\}$. Yet $y_1 \notin \{y^2\}_1$ because $y_1$ appears in $[\{y^2\}_0] = [y^2]$ only in terms which are divisible by $y$ or by 2.

$$\phi \subseteq \{b_1, \cdots, b_s\}_m(^8).$$

One says that *the basis theorem holds in* $\mathfrak{R}$ if every subset of $\mathfrak{R}$ has a basis. If every subset of $\mathfrak{R}$ has an $m$-basis, with $m$ depending on the subset, then we shall say that *the *-basis theorem holds in* $\mathfrak{R}$. If every subset has an $m$-basis, with a single $m$ independent of the subset, we shall say that *the m-basis theorem holds in* $\mathfrak{R}$.

The basis theorem of Ritt and Raudenbush mentioned above is seen to be, in our terminology, a 1-basis theorem.

3. **A useful result.** Let $a$ be an arbitrary element of $\mathfrak{R}$, $\phi$ an arbitrary subset of $\mathfrak{R}$. Denote the set of all elements $af$ ($f \in \phi$) by $a \cdot \phi$.

We shall show that

$$a \cdot \{\phi\}_m \subseteq \{a \cdot \phi\}_m.$$

Indeed, since $\{\phi\}_0 = (\phi)$, the relation in question subsists when $m = 0$. Suppose it holds for $m = k$. Let $f \in \{\phi\}_k$. We show that

$$a^t \delta_1^{i_1} \cdots \delta_r^{i_r} f \in [a \cdot \{\phi\}_k] \subseteq [\{a \cdot \phi\}_k], \qquad t = i_1 + \cdots + i_r + 1.$$

Indeed, since this relation is obvious for $i_1 + \cdots + i_r = 0$ it follows in general from the fact that

$$a^{h+1}\delta_i g = a\delta_i(a^h g) - ha^h g \delta_i a \in [a^h g].$$

Thus, $a \cdot [\{\phi\}_k] \subseteq \{a \cdot \phi\}_{k+1}$. Hence, if $g \in \{\phi\}_{k+1}$, that is, if $g^t \in [\{\phi\}_k]$, then $ag^t \in \{a \cdot \phi\}_{k+1}$, $ag \in \{a \cdot \phi\}_{k+1}$, so that $a \cdot \{\phi\}_{k+1} \subseteq \{a \cdot \phi\}_{k+1}$, q.e.d.

An easy consequence of our result is that

$$\{\phi\}_m \cdot \{\psi\}_n \subseteq \{\phi \cdot \psi\}_{m+n}.$$

4. **Maximal subsets**(9). Let $\mathfrak{M}$ be a collection of subsets of $\mathfrak{R}$ such that every transfinite sequence $\phi_\xi$ of subsets of $\mathfrak{R}$ in $\mathfrak{M}$ which satisfies the condition

$$\phi_\xi \subset \phi_\eta, \qquad \text{if } \xi < \eta,$$

also satisfies the condition

$$\Sigma \phi_\xi \in \mathfrak{M}.$$

We shall prove that $\mathfrak{M}$ *contains a maximal subset of* $\mathfrak{R}$, that is, a $\phi \in \mathfrak{M}$ such that $\psi \in \mathfrak{M}$ implies $\phi \not\subset \psi$.

Indeed, let $\phi_\xi$ be a well-ordering of $\mathfrak{M}$. Define by transfinite induction:

$\psi_1 = \phi_1$,

$\psi_\eta =$ the first $\phi_\xi$ such that $\psi_\nu \subset \phi_\xi$ for all $\nu < \eta$.

By the construction, no $\phi_\eta$ properly contains every $\psi_\eta$. The resulting transfinite sequence $\psi_\eta$ must have a last element. For otherwise $\Sigma\psi_\eta$ would be a $\phi_\xi$ properly containing every $\psi_\eta$. This last element is a maximal subset.

(8) Thus, if $m \leqq n$, every $m$-basis is an $n$-basis.

(9) In this section $\mathfrak{R}$ may be an abstract set.

5. **Systems of differential polynomials of bounded order.** We suppose henceforth that $\mathfrak{R}$ contains a unit element 1.

Let $y_1, \cdots, y_n$ be unknowns, and let $\Phi$ be a set of (partial) differential polynomials, or forms, in $\mathfrak{R}\{y_1, \cdots, y_n\}$ [10] of bounded orders. We shall show that *if the basis (or *-basis) theorem holds in $\mathfrak{R}$ then $\Phi$ has a basis (or $m$-basis, for some finite $m$).*

**Proof.** Because the differential polynomials in $\Phi$ are of bounded orders, only a finite number of partial derivatives of the $y_i$ are effectively present in the forms of $\Phi$. Let $q$ be the least integer such that there exists a set $\Phi$, involving only $q$ derivatives of the $y_i$ which has no basis (or $m$-basis). By the hypothesis on $\mathfrak{R}$, $q>0$. We work toward a contradiction.

If $\Phi_\xi$ is a transfinite sequence of sets of differential polynomials in $\mathfrak{R}\{y_1, \cdots, y_n\}$ involving only $q$ derivatives of the $y_i$ such that

$$\Phi_\xi \subset \Phi_\eta, \qquad \text{if } \xi < \eta,$$

no $\Phi_\xi$ having a basis (or $m$-basis), then $\Sigma\Phi_\xi$ involves only $q$ derivatives of the $y_i$ and has no basis (or $m$-basis). For if $\Sigma\Phi_\xi$ had a basis (or $m$-basis) there would be a single $\Phi_\xi$ which would contain every differential polynomial of the basis, and that $\Phi_\xi$ itself would have a basis (or $m$-basis). Therefore, by §4, there is a maximal set of forms involving only $q$ derivatives of the $y_i$ which has no basis (or $m$-basis).

Let $\Phi$ be such a maximal set. Denote the $q$ partial derivatives of the $y_i$ present in $\Phi$ by $\alpha_1, \cdots, \alpha_q$.

It is clear that $\Phi$ is an ideal in $\mathfrak{R}[\alpha_1, \cdots, \alpha_q]$, for otherwise the ideal generated by $\Phi$ in $\mathfrak{R}[\alpha_1, \cdots, \alpha_q]$ would properly contain $\Phi$, would involve only $q$ derivatives, and would have no basis (or $m$-basis).

Let $\Phi'$ be the set of differential polynomials in $\Phi$ which are free of $\alpha_q$.

If every element of $\Phi$, written as a polynomial in $\alpha_q$, had each coefficient in $\Phi'$, we would have $\Phi\subseteq(\Phi')$, so that $\Phi$ would have a basis (or $m$-basis), because $\Phi'$ does. Hence $\Phi$ contains a form in which $\alpha_q$ is effectively present and which, when written as a polynomial in $\alpha_q$, has its leading coefficient not in $\Phi$.

Of all such differential polynomials let

$$B = I\alpha_q^s + \cdots, \qquad I \notin \Phi,$$

be one of minimum degree $s$ in $\alpha_q$. Then, for each $G\in\Phi$, we have, for suitable $t$,

$$I^tG \equiv G' \ (B),$$

where $G'\in\Phi$ has its degree in $\alpha_q$ less than $s$[11]. By the minimal nature of

[10] $\mathfrak{R}\{y_1, \cdots, y_n\}$ means the ring obtained by the differential ring adjunction of $y_1, \cdots, y_n$ to $\mathfrak{R}$.

[11] Here we use for the first time the fact that $\mathfrak{R}$ contains a unit element.

the degree of $B$ it follows that $G' \in (\Phi')$. But $\Phi'$ has a basis (or $m_1$-basis) $D_1, \cdots, D_u$. Hence $IG \in \{B, D_1, \cdots, D_u\}$ (or $IG \in \{B, D_1, \cdots, D_u\}_{m_1}$).

Now, by the maximality of $\Phi$, $(I, \Phi)$ has a basis (or $m_2$-basis) which we may write as $I, D_{u+1}, \cdots, D_v$, where each $D_i \in \Phi$. Hence, referring to §3,

$$\begin{aligned} G^2 &\in G\cdot(I, \Phi) \subseteq G\cdot\{I, D_{u+1}, \cdots, D_v\} \\ &\subseteq \{IG, D_{u+1}, \cdots, D_v\} \subseteq \{\{B, D_1, \cdots, D_u\}, D_{u+1}, \cdots, D_v\}, \\ G &\in \{B, D_1, \cdots, D_v\}, \\ \Phi &\subseteq \{B, D_1, \cdots, D_v\} \end{aligned}$$

(or, similarly, $\Phi \subseteq \{B, D_1, \cdots, D_v\}_{m_1+m_2}$). This contradiction completes the proof.

6. **Regular differential rings.** A differential ring $\mathfrak{R}$ will be called *regular* if every prime differential ideal $\pi \subseteq \mathfrak{R}$ which contains a prime rational integer $p$ is such that the congruence

$$a \equiv x^p\ (\pi)$$

has a solution $x \in \mathfrak{R}$ for every $a \in \mathfrak{R}$ (that is, if every element has a $p$th root modulo $\pi$).

If $\mathfrak{R}$ is of characteristic $p > 0$ then every ideal contains $p$ and no ideal other than $\mathfrak{R}$ itself contains a prime number different from $p$.

Examples of regular differential rings are:

1. every differential ring which contains the rational number system;
2. every differential ring with unit element of characteristic $q > 0$ in which each element has a $q$th root;
3. every perfect ("vollständig") differential field;
4. the differential ring of rational integers.

7. **The basis theorem.** The theorem we shall prove is the following:

*Let $\mathfrak{R}$ be a regular commutative differential ring with unit element. Let $\mathfrak{R}'$ be a commutative differential ring obtained from $\mathfrak{R}$ by the differential ring adjunction of a finite number of elements: $\mathfrak{R}' = \mathfrak{R}\{\eta_1, \cdots, \eta_n\}$*[12]. *If the basis (or *-basis) theorem holds in $\mathfrak{R}$ then the basis (or *-basis) theorem holds in $\mathfrak{R}'$.*

It is necessary to prove the theorem only for the case in which the $\eta_i$ are all unknowns, $\eta_i = y_i$; for if the basis theorem holds in $\mathfrak{R}\{y_1, \cdots, y_n\}$ then it is easy to see that it will continue to hold when any or all of the $y_i$ are replaced by elements among which an algebraic differential relation subsists.

8. **The proof begun.** Assume that there exists in $\mathfrak{R}' = \mathfrak{R}\{y_1, \cdots, y_n\}$ a system which does not have a basis (or $m$-basis for any $m$).

If $\Sigma_\xi$ is a transfinite sequence of such systems with $\Sigma_\xi \subset \Sigma_\eta$ whenever $\xi < \eta$ then the logical sum of the $\Sigma_\xi$ is again such a system. For if the logical

[12] The $\eta_i$ may be hypertranscendental over $\mathfrak{R}$ (for example, they may be unknowns) or may satisfy some algebraic differential relation with coefficients in $\mathfrak{R}$.

sum had a basis (or $m$-basis) then there would be a single $\Sigma_\xi$ which would contain every form of the basis, and that $\Sigma_\xi$ itself would have a basis (or $m$-basis).

By §4 it follows that there is a maximal system which has no basis (or $m$-basis). We let $\Sigma$ be such a maximal system and seek a contradiction.

$\Sigma$ is a differential ideal, for $[\Sigma]$, like $\Sigma$, has no basis (or $m$-basis) and therefore can not properly contain $\Sigma$. Moreover, $\Sigma$ is prime. To prove this, assume to the contrary that $AB\in\Sigma$, $A\notin\Sigma$, $B\notin\Sigma$. Then $(\Sigma, A)$ and $(\Sigma, B)$ properly contain $\Sigma$ and must have bases (or $m_1$- and $m_2$-bases, respectively), say $A, C_1, \cdots, C_u$ and $B, C_{u+1}, \cdots, C_v$, respectively, where the $C_i$ are in $\Sigma$. Thus

$$\Sigma^2 \subseteq (\Sigma, A)(\Sigma, B) \subseteq \{A, C_1, \cdots, C_u\}\{B, C_{u+1}, \cdots, C_v\} \subseteq \{AB, C_1, \cdots, C_v\},$$

so that $\Sigma\subseteq\{AB, C_1, \cdots, C_v\}$, and $\Sigma$ has a basis (or, similarly, an $(m_1+m_2)$-basis).

9. **The proof continued.** The object of this section is to show that $\Sigma$ contains a prime rational integer $p$(13). To accomplish this we introduce a set of differential polynomials analogous to the "basic sets" used by Ritt.

We assume that the partial derivatives of the $y_i$ are completely ordered by a system of marks in such a way that every partial derivative of the $y_i$ is lower than (precedes) every other derivative of the $y_i$ of higher order, and if $\alpha$ and $\beta$ are two derivatives of the $y_i$ with $\alpha$ lower than $\beta$ then $\delta_i\alpha$ is lower than $\delta_i\beta$, $i=1, \cdots, r$. Such an ordering can always be effected(14).

Let $\sigma=\Sigma\cap\mathfrak{R}$. Clearly $\sigma$ is a prime differential ideal in $\mathfrak{R}$. Since the basis (or $*$-basis) theorem holds in $\mathfrak{R}$, $\sigma$ has a basis (or $m$-basis). Hence $\Sigma\neq(\sigma)$ so that $\Sigma$ must contain forms none of whose coefficients is in $\sigma$.

Of all the forms in $\Sigma$ none of whose coefficients is in $\sigma$, consider those with lowest possible leader $\alpha_1$ (the leader of a form is the highest derivative of the $y_i$ effectively present in the form). Of all those forms let $A_1$ be one whose degree in $\alpha_1$ is as low as possible.

Of all the forms in $\Sigma$ none of whose coefficients is in $\sigma$, which do not contain a proper derivative (that is, a derivative of positive order) of $\alpha_1$, and which are of lower degree in $\alpha_1$ than $A_1$, consider those with lowest possible leader $\alpha_2$. Of all those let $A_2$ be one whose degree in $\alpha_2$ is a minimum.

Continuing, at the $j$th step, consider, of the forms in $\Sigma$ none of whose coefficients is in $\sigma$, which do not contain a proper derivative of $\alpha_i$ $(i=1, \cdots, j-1)$ and which are of lower degree in $\alpha_i$ than $A_i$ $(i=1, \cdots, j-1)$, those forms which have the lowest leader $\alpha_j$. Of all those forms let $A_j$ be one whose degree in $\alpha_j$ is a minimum.

Since no $\alpha_i$ is a derivative of any preceding $\alpha_i$, there can be only a finite

(13) If $\mathfrak{R}$ contained all the rational numbers this would suffice, for then $\Sigma$ would contain $1=(1/p)\cdot p$, and would have 1 as a basis.

(14) Ritt, loc. cit., pp. 141–143.

number of the $\alpha_i$[15], so that the process for defining the forms $A_i$ must stop. Let $A_s$ be the last $A_i$.

It is easy to see that if $G$ is a form in $\Sigma$ which contains no proper derivative of any $\alpha_i$ and whose degree in each $\alpha_i$ is lower than that of the corresponding $A_i$, then $G\in(\sigma)$.

Let $I_i$ and $S_i$ be the initial and separant of $A_i$. The coefficients of $I_i$ are coefficients of $A_i$ and therefore are not in $\sigma$. $I_i$ contains no proper derivative of any $\alpha_j$ and is of lower degree in $\alpha_j$ than $A_j$ $(j=1, \cdots, s)$. Hence $I_i \notin \Sigma$.

We shall show that at least one $S_i$ is in $\Sigma$.

Let no $S_i$ be in $\Sigma$. For an arbitrary form $G\in\Sigma$ there exist integers $g_i$, $h_i$ such that

$$I_1^{g_1}S_1^{h_1}\cdots I_s^{g_s}S_s^{h_s}G \equiv G'[A_1, \cdots, A_s],$$

where $G'\in\Sigma$ contains no proper derivative of any $\alpha_i$ and is of lower degree in $\alpha_i$ than $A_i$. Thus, by the above, $G'\in(\sigma)$, so that

$$I_1^{g_1}S_1^{h_1}\cdots I_s^{g_s}S_s^{h_s}G \equiv 0[A_1, \cdots, A_s, \sigma],$$

$$I_1S_1\cdots I_sS_sG \in \{A_1, \cdots, A_s, \sigma\}_1,$$

$$I_1S_1\cdots I_sS_s\cdot\Sigma \subseteq \{A_1, \cdots, A_s, \sigma\}_1.$$

Now, $\Sigma$ is prime and contains no $I_i$ or $S_i$, so that $I_1S_1\cdots I_sS_s \notin \Sigma$. Hence, by the maximality of $\Sigma$, the system

$$\Sigma,\ I_1S_1\cdots I_sS_s$$

has a basis (or $m_1$-basis) which we may write as

$$I_1S_1\cdots I_sS_s,\ B_1, \cdots, B_t.$$

Denoting by $B_{t+1}, \cdots, B_u$ a basis (or $m_2$-basis) of $\sigma$, we have

$$\begin{aligned}\Sigma^2 &\subseteq \Sigma(\Sigma, I_1S_1\cdots I_sS_s) \subseteq \Sigma\{I_1S_1\cdots I_sS_s, B_1, \cdots, B_t\}\\ &\subseteq \{I_1S_1\cdots I_sS_s\cdot\Sigma, B_1, \cdots, B_t\} \subseteq \{A_1, \cdots, A_s, \sigma, B_1, \cdots, B_t\}\\ &\subseteq \{A_1, \cdots, A_s, B_1, \cdots, B_u\},\\ \Sigma &\subseteq \{A_1, \cdots, A_s, B_1, \cdots, B_u\}\end{aligned}$$

(or, similarly, $\Sigma\subseteq\{A_1, \cdots, A_s, B_1, \cdots, B_u\}_{m_1+m_2+1}$). This contradicts the fact that $\Sigma$ has no basis (or $m$-basis) and proves that $S_i\in\Sigma$ for at least one $i$.

Let $S_j$ be the first $S_i$ contained in $\Sigma$. $S_j$ contains no proper derivative of any $\alpha_i$ and is of lower degree in $\alpha_i$ than $A_i$ $(i=1, \cdots, s)$. Hence $S_j\in(\sigma)$. It follows that $n_jI_j\in\Sigma$, where $n_j$ is the degree of $A_j$ in $\alpha_j$, so that $n_j\in\Sigma$, and one of the prime factors $p$ of $n_j$ must be in $\Sigma$. This completes the proof of the result at the beginning of this section.

(15) Ritt, loc. cit., pp. 135–136.

10. **The proof concluded.** Let $F$ be any nonzero differential polynomial, $\gamma$ any partial derivative of the $y_i$. $F$ can be written in one and only one way as a polynomial

$$H_0 + H_1\gamma + \cdots + H_h\gamma^h, \qquad H_h \neq 0,$$

in $\gamma$ of degree $h < p$, where $\gamma$ does not appear in the $H_i$ except raised to powers divisible by $p$. We shall call $h$ the $p$-degree of $F$ in $\gamma$. The highest derivative of the $y_i$ in which $F$ has a positive $p$-degree (if such a derivative exists) shall be called the $p$-leader of $F$. If $\gamma$ is the $p$-leader of $F$ and if $h$ is the $p$-degree of $F$ in $\gamma$, we shall call the coefficient of $\gamma^h$ the $p$-initial of $F$, and $\partial F/\partial\gamma$ the $p$-separant of $F$.

We shall need the fact that $\Sigma$ contains a form, none of whose coefficients is in $\sigma$, whose $p$-degree in some derivative of the $y_i$ is positive and whose $p$-initial is not in $\Sigma$. To prove this assume the contrary and let $G$ be a form of $\Sigma$, none of whose coefficients is in $\sigma$, of least possible (total) degree. Every term of $G$ involves only powers divisible by $p$, else the $p$-degree of $G$ in some derivative of the $y_i$ would be positive and the $p$-initial of $G$ would be a form in $\Sigma$, none of whose coefficients is in $\sigma$, of lower degree than $G$. Moreover, by the regularity of $\mathfrak{R}$, the coefficient of each term of $G$ may be replaced modulo $\sigma$ by the $p$th power of an element of $\mathfrak{R}$([16]). Since $p \in \sigma$ it follows that $G \equiv H^p\ (\sigma)$, where $H$ is the form obtained from $G$ by replacing each term by its $p$th root modulo $\sigma$. $H$ is of lower degree than $G$ and is in $\Sigma$. This contradicts the definition of $G$ and proves the required fact.

Of all the forms of $\Sigma$, none of whose coefficients is in $\sigma$, which involve derivatives of the $y_i$ to a power not divisible by $p$ and whose $p$-initials are not in $\Sigma$, consider those with lowest $p$-leader $\beta_1$. Of all those forms let $B_1$ be one whose $p$-degree in $\beta_1$ is as low as possible.

Of all the forms in $\Sigma$, none of whose coefficients is in $\sigma$, which involve derivatives of the $y_i$ to a power not a multiple of $p$, whose $p$-initials are not in $\Sigma$, which do not contain a proper derivative of $\beta_1$ except raised to a power divisible by $p$, and which have a $p$-degree in $\beta_1$ less than that of $B_1$, consider those with lowest possible $p$-leader $\beta_2$. Of all those let $B_2$ be one whose $p$-degree in $\beta_2$ is a minimum.

Continuing, at the $j$th step, of all the forms in $\Sigma$, none of whose coefficients is in $\sigma$, which contain derivatives of the $y_i$ to powers not divisible by $p$, whose $p$-initials are not in $\Sigma$, which do not contain a proper derivative of $\beta_i$ except to a power divisible by $p$ $(i=1, \cdots, j-1)$ and which have a $p$-degree in $\beta_i$ less than that of $B_i$ $(i=1, \cdots, j-1)$, consider those with lowest $p$-leader $\beta_j$. Of all those forms let $B_j$ be one whose $p$-degree in $\beta_j$ is as low as possible.

As with the $A_i$ of §9, the process of defining the $B_i$ must stop after a finite number of steps. Let $B_s$ be the last $B_i$([17]). Let $J_i$ and $T_i$ be the $p$-initial and $p$-separant, respectively, of $B_i$.

([16]) Up to this point we have not used the regularity. Henceforth it will be important.

([17]) The $s$ here is not necessarily the same as that of §9.

If $G$ is a form of $\Sigma$, none of whose coefficients is in $\sigma$, which contains no proper derivative of any $\beta_i$ except raised to powers divisible by $p$ and whose $p$-degree in each $\beta_i$ is less than that of the corresponding $B_i$, then either $G$ contains no derivative of the $y_i$ that is raised to a power not divisible by $p$, or the $p$-initial of $G$ is in $\Sigma$.

From this it can be shown that the $T_i$ are not contained in $\Sigma$. We already know that the $I_i$ are not in $\Sigma$.

Let $\alpha$ represent the highest derivative of the $y_i$ effectively present in $B_1, \cdots, B_s$([18]). Let $\Sigma_\alpha$ denote the totality of forms in $\Sigma$ which contain no derivative of the $y_i$ which is higher than $\alpha$. The forms of $\Sigma_\alpha$ are of bounded order.

We shall show that for each differential polynomial $G \in \Sigma$ there exist non-negative integers $e_i, f_i$ such that

$$J_1^{e_1} T_1^{f_1} \cdots J_s^{e_s} T_s^{f_s} G \equiv 0[\Sigma_\alpha].$$

Assume that this is not so. If $G$ is a form in $\Sigma$ for which such a congruence fails to hold it is easy to see that there is a relation

$$J_1^{g_1} T_1^{h_1} \cdots J_s^{g_s} T_s^{h_s} G \equiv G'[B_1, \cdots, B_s],$$

where $G'$ is a form in $\Sigma$ for which such a congruence fails to hold, which contains no proper derivative of any $\beta_i$ except to a power divisible by $p$, and which has, in each $\beta_i$, a $p$-degree lower than that of the corresponding $B_i$. Of all forms in $\Sigma$ which fail to satisfy a congruence as above, which contain no proper derivative of any $\beta_i$ except to a power divisible $p$, and which have, in each $\beta_i$, a $p$-degree lower than that of the corresponding $B_i$, consider those with the least number of terms. Of all those forms let $G$ be one with a minimum (total) degree. Since $\sigma \subseteq \Sigma_\alpha$, no coefficient of $G$ is in $\sigma$. Hence, either $G$ contains only powers divisible by $p$ or the $p$-initial of $G$ is in $\Sigma$. Suppose $G$ contains only powers divisible by $p$. By the regularity of $\mathfrak{R}$, each coefficient of $G$ may be replaced modulo $\sigma$ by the $p$th power of an element of $\mathfrak{R}$. Hence $G \equiv H^p$ $(\sigma)$, where $H \in \Sigma$, having the same number of terms as $G$ and being of lower degree than $G$, satisfies a congruence as above. But this is impossible as then $G$ would satisfy such a congruence. Thus $G$ has a $p$-leader $\gamma$ and the $p$-initial of $G$ is in $\Sigma$:

$$G = K_0 + \cdots + K_g \gamma^g, \qquad K_g \in \Sigma.$$

Since $K_g$ is of lower degree than $G$, $K_g$ satisfies a congruence of the type in question. But $K_0 + \cdots + K_{g-1}\gamma^{g-1}$, which is in $\Sigma$ and has fewer terms than $G$, must also satisfy such a congruence. This is impossible, however, for it implies that $G$ itself satisfies the same kind of congruence.

Thus we have shown that

([18]) $\alpha$ may be higher than $\beta_s$ as the $p$-degree of each $B_i$ in $\alpha$ may be 0.

$$J_1T_1 \cdots J_sT_s \cdot \Sigma \subseteq \{\Sigma_\alpha\}_1.$$

Now, by the result of §5, $\Sigma_\alpha$ has a basis (or $m_1$-basis), say $D_1, \cdots, D_t$. Also, by the maximality of $\Sigma$, the system

$$\Sigma,\ J_1T_1 \cdots J_sT_s$$

has a basis (or $m_2$-basis) which we may write as

$$J_1T_1 \cdots J_sT_s,\ D_{t+1}, \cdots, D_u \qquad D_i \in \Sigma.$$

Hence, by §3,

$$\begin{aligned}\Sigma^2 &\subseteq \Sigma(\Sigma, J_1T_1 \cdots J_sT_s) \subseteq \Sigma\{J_1T_1 \cdots J_sT_s, D_{t+1}, \cdots, D_u\}\\ &\subseteq \{J_1T_1 \cdots J_sT_s \cdot \Sigma, D_{t+1}, \cdots, D_u\} \subseteq \{D_1, \cdots, D_u\},\\ \Sigma &\subseteq \{D_1, \cdots, D_u\}\end{aligned}$$

(or, similarly, $\Sigma \subseteq \{D_1, \cdots, D_u\}_{m_1+m_2+1}$). This contradiction completes the proof of the theorem stated in §7.

11. **Shorter proof in the ordinary case.** We sketch in this section a shorter proof of the theorem under the assumption that we are dealing with *ordinary* differential rings, that is, differential rings with one type of differentiation.

Denote the $j$th derivative of any letter $u$ by $u_j$.

Of the above proof we take over §§1–6.

We first show that, when the basis (or *-basis) theorem holds in $\mathfrak{R}$ and $\mathfrak{R}$ is regular, the basis (or *-basis) theorem holds in $\mathfrak{R}\{y\}$. Assuming the contrary we obtain, as in §8, a maximal system $\Sigma \subset \mathfrak{R}\{y\}$ which has no basis (or $m$-basis). $\Sigma$ is a prime differential ideal. If $F$ were a form in $\Sigma$ whose separant $S$ was not in $\Sigma$, $S \cdot \Sigma$ would have a basis (or $m_1$-basis), $B_1, \cdots, B_s$, for which we could write, for each $G \in \Sigma$, $S^pG \equiv G'[F]$, with $G'$ of order no higher than that of $F$, so that we would have $S \cdot \Sigma \subseteq \{\Sigma'\}$, where $\Sigma'$ is the set of forms of $\Sigma$ whose orders are less than or equal to the order of $F$. Also, by the maximality of $\Sigma$, the system $\Sigma$, $S$ would have a basis (or $m_2$-basis), say $S, B_{s+1}, \cdots, B_t$. Thus we would have

$$\begin{aligned}\Sigma^2 &\subseteq \Sigma(\Sigma, S) \subseteq \Sigma\{S, B_{s+1}, \cdots, B_t\}\\ &\subseteq \{S \cdot \Sigma, B_{s+1}, \cdots, B_t\} \subseteq \{B_1, \cdots, B_t\},\\ \Sigma &\subseteq \{B_1, \cdots, B_t\}\end{aligned}$$

(or, similarly, $\Sigma \subseteq \{B_1, \cdots, B_t\}_{m_1+m_2}$). This cannot be, so that every form in $\Sigma$ must have its separant in $\Sigma$.

Of all forms in $\Sigma$ none of whose coefficients is in $\Sigma$ let $A$ be one whose (total) degree is a minimum. Since $S$, the separant of $A$, is of lower degree than $A$, all the coefficients of $S$ must be in $\Sigma$. These coefficients are coefficients of $A$ multiplied by the exponents to which $y_q$ appears in $A$. (Here $q$ is the order of $A$.) Since $\Sigma$ is prime and the coefficients of $A$ are not in $\Sigma$,

these exponents must be in $\Sigma$. These exponents have a common prime factor $p \in \Sigma$, and we see that $y_q$ appears in $A$ only to powers divisible by $p$. It is now easy to see that every derivative of $y$ appears in $A$ only to powers divisible by $p$; for suppose $y_j$ is the $y_i$ of highest subscript which appears in $A$ to a power not a multiple of $p$. Then the $(q-j+1)$st derivative of $A$ would be, terms divisible by $p$ neglected, a form in $\Sigma$ whose separant is not in $\Sigma$, an impossibility.

Now, by the regularity of $\mathfrak{R}$, we may replace modulo $\Sigma$ each coefficient of $A$ by the $p$th power of an element of $\mathfrak{R}$. Hence $A \equiv B^p(\Sigma)$, where $B \in \Sigma$ has no coefficient in $\Sigma$ and is of lower degree than $A$. This completes the proof for $\mathfrak{R}\{y\}$.

Proceeding by induction, suppose the theorem has been proved for $\mathfrak{R}\{y_1, \cdots, y_{n-1}\}$(19). As above, we find, for a maximal system $\Sigma \subset \mathfrak{R}\{y_1, \cdots, y_n\}$, that the separant of each form of $\Sigma$ must itself be in $\Sigma$. This must be true no matter how we order the unknowns. Letting $A$ be a form in $\Sigma$, with no coefficients in $\Sigma$, of minimum degree, we see from the above that each $y_{ij}$ appears in $A$ only to powers divisible by a prime rational integer $p \in \Sigma$. As in the case of one unknown this leads to a contradiction and completes the proof.

12. **Examples.** From the point of view of analogy with the Hilbert basis theorem, it might be imagined that the regularity condition imposed in the basis theorem above is unnecessary. The following example shows that this is not so.

EXAMPLE 1. Let $\mathfrak{R}$ be the ordinary differential field of characteristic $p>0$ obtained from the field of rational integers modulo $p$ by the differential field adjunction of the set of "indeterminate constants" $c_0, c_1, c_2, \cdots$, that is, each $c_i$ is a letter whose derivative is taken to be 0, and $\mathfrak{R} = \mathfrak{J}_p\langle c_0, c_1, c_2, \cdots \rangle$. Let $y$ be an unknown and consider, in $\mathfrak{R}\{y\}$, the system $\Phi$:

$$y^p + c_0,\ y_1^p + c_1, \cdots,\ y_k^p + c_k, \cdots.$$

We shall show that $\Phi$ *has no basis.*

Indeed, if $\Phi$ had a basis we should have, for some $k$,

$$y_k^p + c_k \in \{y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}\}.$$

Now, $(y_i^p+c_i)_1 = py_i^{p-1}y_{i+1}+c_{i1}=0$, so that

$$[y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}] = (y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}).$$

But clearly $AB \in (y^p+c_0, \cdots, y_{k-1}^p+c_{k-1})$ implies that $A$ or $B \in (y^p+c_0, \cdots, y_{k-1}^p+c_{k-1})$. Hence $(y^p+c_0, \cdots, y_{k-1}^p+c_{k-1})$ is a prime differential ideal, so that

$$\{y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}\} = (y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}).$$

(19) The $y_i$ are unknowns. The $j$th derivative of $y_i$ is denoted by $y_{ij}$.

But it is easy to verify that

$$y_k^p + c_k \notin (y^p + c_0, \cdots, y_{k-1}^p + c_{k-1}).$$

Let $\mathfrak{J}$ be the ordinary differential ring of rational integers. Let $n = 2^{m-1}$, where $m$ is any positive integer. The following example shows that the $m$-basis theorem does not hold in $\mathfrak{J}\{y_1, \cdots, y_n\}$.

EXAMPLE 2. Let $\Phi$ be the system in $\mathfrak{J}\{y_1, \cdots, y_n\}$ consisting of the forms

$$y_1^2 \cdots y_n^2, \; y_{11}^2 \cdots y_{n1}^2, \cdots, y_{1k}^2 \cdots y_{nk}^2, \cdots .$$

*$\Phi$ has no $m$-basis.*

To prove this assume the contrary. Then, for some $k$,

$$\begin{aligned} y_{1k}^2 \cdots y_{n'k}^2 &\in \{y_1^2 \cdots y_{n'}^2, \cdots, y_{1,k-1}^2 \cdots y_{n',k-1}^2\}_m \\ &\subseteq \{2, y_1 \cdots y_{n'}, \cdots, y_{1,k-1} \cdots y_{n',k-1}\}_{m-1}, \\ y_{1k} \cdots y_{n'k} &\in \{2, y_1 \cdots y_{n'}, \cdots, y_{1,k-1} \cdots y_{n',k-1}\}_{m-1}. \end{aligned}$$

Letting $y_1 = y_2 = z_1$, $y_3 = y_4 = z_2, \cdots, y_{n-1} = y_n = z_{n'}$, where $n' = n/2 = 2^{m-2}$, we see, in the differential ring $\mathfrak{R}\{z_1, \cdots, z_{n'}\}$, that

$$\begin{aligned} z_{1k}^2 \cdots z_{n'k}^2 &\in \{2, z_1^2 \cdots z_{n'}^2, \cdots, z_{1,k-1}^2 \cdots z_{n',k-1}^2\}_{m-1} \\ &\subseteq \{2, z_1 \cdots z_{n'}, \cdots, z_{1,k-1} \cdots z_{n',k-1}\}_{m-2}, \\ z_{1k} \cdots z_{n'k} &\in \{2, z_1 \cdots z_{n'}, \cdots, z_{1,k-1} \cdots z_{n',k-1}\}_{m-2}. \end{aligned}$$

Continuing, at each step we reduce the number of unknowns by one half until we arrive, in $\mathfrak{R}\{u_1\}$, at the relation

$$u_{1k} \in \{2, u_1, \cdots, u_{1,k-1}\}_0 = (2, u_1, \cdots, u_{1,k-1}).$$

This contradiction completes the proof.

COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# ON THE DERIVATIVES OF FUNCTIONS ANALYTIC IN THE UNIT CIRCLE AND THEIR RADII OF UNIVALENCE AND OF $p$-VALENCE

BY

W. SEIDEL AND J. L. WALSH

## TABLE OF CONTENTS



---

Presented to the Society, February 26, 1938; received by the editors July 7, 1941. A short abstract of this paper under a slightly different title without details of proofs has already been published, Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 337–340.

CHAPTER V. MISCELLANEOUS



1. **Introduction.** Various results are known concerning the order of growth of the first and higher derivatives of univalent and of bounded functions analytic in the unit circle, in the plane of the complex variable $z$. Among these may be mentioned Koebe's distortion theorem (Verzerrungssatz) in the univalent case, and Schwarz's lemma and the results of O. Szász(1) in the bounded case. A consequence of these results for a function $f(z)$ analytic in $|z|<1$ is $|f'(z)|=O((1-|z|)^{-3})$ in the case that $f(z)$ is univalent and $|f^{(n)}(z)|=O((1-|z|)^{-n})$ in the case that $f(z)$ is bounded. Various distortion theorems for bounded univalent functions were found by G. Pick and R. Nevanlinna(2). H. Frazer and more recently M. L. Cartwright have obtained results on the order of growth of $p$-valent functions(3) in a complete form.

All these investigations, however, fail to give an adequate description of the behavior of $|f'(z)|\ (1-|z|)$ as $|z|\to 1$ from the interior of the unit circle $|z|<1$. In the univalent case an answer to this question is contained in the following result due to J. E. Littlewood without the precise constant involved and to A. J. Macintyre(4) in the precise form stated here.

THEOREM 1. *Let $f(z)$ be analytic and univalent in $|z|<1$ and let it omit there the value $\omega$. Then, in $|z|<1$ the following inequality is satisfied:*

$$|f'(z)|\,(1-|z|^2)\leqq 4\,|\omega-f(z)|. \tag{1.1}$$

Theorem 1 is in fact essentially one form of Koebe's distortion theorem, as we indicate below.

The object of the present paper is to study in some detail the behavior of expressions of the form $|f^{(p)}(z)|\ (1-|z|)^p$ for various classes of functions $f(z)$ analytic in the unit circle $|z|<1$, especially the behavior as $|z|\to 1$. We thus obtain results which can be interpreted as new distortion theorems. In

(1) O. Szász, Mathematische Zeitschrift, vol. 8 (1920), pp. 303–309.

(2) G. Pick, Sitzungsberichte der Kaiserlichen Akademie der Wissenschaften, Vienna, Abteilung IIa, vol. 126 (1917), pp. 247–263; R. Nevanlinna, Översigt af Finska Vetenskaps Societetens Förhandlingar, vol. 62 (1919).

(3) M. L. Cartwright, Mathematische Annalen, vol. 111 (1935), pp. 98–118.

(4) J. E. Littlewood, Proceedings of the London Mathematical Society, vol. 23 (1924) p. 507; A. J. Macintyre, Journal of the London Mathematical Society, vol. 11 (1936), pp. 7–11.

particular, the expression $|f'(z)|\ (1-|z|^2)$ is found to be closely connected with the radius of univalence, which is now to be defined.

DEFINITION 1. *Let $w=f(z)$ be analytic in $|z|<1$ and let $R$ denote the Riemann configuration*(5) *over the $w$-plane onto which this function maps the region $|z|<1$. Let $w_0$ be an arbitrary point, not a branch point, of $R$. Then the radius of the largest smooth circle (boundary not included) with center at $w_0$ and wholly contained in $R$ is called the radius of univalence of $R$ at $w_0$ and will be denoted by $D_1(w_0)$. At a branch point $w_0$ of $R$ we define $D_1(w_0)$ as zero.*

In this definition $w_0$ refers to an actual point of $R$ and not merely to any point of $R$ whose affix is the complex number $w_0$; the notation $D_1(w_0)$ is thus not fully explicit. The reader will easily verify that the largest smooth circle whose existence is asserted in the definition does exist and is unique.

This terminology differs from that of Montel(6), who uses the term modulus of univalence for our radius of univalence. A similar comment applies to the terminology radius of $p$-valence which we define in §14.

Explicit inequalities connecting $|f'(z)|\ (1-|z|^2)$ and $D_1(w)$ are obtained for the class of functions $f(z)$ univalent in $|z|<1$ in Theorem 3, Chapter I, for functions $f(z)$ bounded in $|z|<1$ in Theorem 3, Chapter II, and for functions $f(z)$ omitting two values in $|z|<1$ in Theorems 2 and 4 of Chapter IV. Analogous to the inequalities connecting $|f'(z)|\ (1-|z|^2)$ and $D_1(w)$ we determine inequalities connecting $|f^{(k)}(z)|\ (1-|z|^2)^k$ for $k=1, 2, \cdots, p$ and $D_p(w)$, where $D_p(w)$ is the radius of the largest $p$-sheeted circle with center in the point $w$ contained in $R$. For the precise definitions the reader may be referred to Chapter II, §§13, 14. We obtain such inequalities on higher derivatives for the class of univalent functions in Theorem 5, Chapter I, for bounded functions in Theorems 1 and 2 of Chapter III, and for the functions omitting two values in Theorem 5, Chapter IV. For the detailed analysis of the paper the reader is referred to the Table of Contents.

Applications of the results just mentioned occur throughout the paper, particularly in Chapter V.

## CHAPTER I. UNIVALENT FUNCTIONS

2. **Preliminary identities.** In the sequel we shall make extensive use of a lemma due to O. Szász(7).

LEMMA 1. *Let $f(z)$ be a function analytic in the circle $|z|<1$. Let*

---

(5) We use the term *Riemann configuration* on which the function $w=f(z)$ regular in $|z|<1$ maps the circle $|z|<1$ to denote that subregion of the Riemann surface of the inverse function of $w=f(z)$ which corresponds to the circle $|z|<1$.

(6) *Leçons sur les Fonctions Univalentes ou Multivalents*, Paris, 1933, pp. 22 and 110.

(7) O. Szász, Mathematische Zeitschrift, vol. 8 (1920), pp. 306–307.

$$g(\zeta) = f\left(\frac{\zeta + z}{1 + \bar{z}\zeta}\right). \tag{2.1}$$

*Then $g(\zeta)$ is a function regular in $|\zeta| < 1$ for every value of $z$ in $|z| < 1$ and*

$$\frac{(1 - |z|^2)^n}{n!} f^{(n)}(z) = \frac{g^{(n)}(0)}{n!} + C_{n-1,1}\bar{z}\frac{g^{(n-1)}(0)}{(n-1)!} + C_{n-1,2}\bar{z}^2\frac{g^{(n-2)}(0)}{(n-2)!} + \cdots + \bar{z}^{n-1}g'(0). \tag{2.2}$$

We omit the proof of Lemma 1 and proceed to the proof of

LEMMA 2. *Let $f(z)$ be a function analytic in the circle $|z| < 1$. Let*

$$g(\zeta) = f\left(\frac{\zeta + z}{1 + \bar{z}\zeta}\right).$$

*Then, for every fixed value of $z$ in $|z| < 1$, $g(\zeta)$ is a function of $\zeta$ regular in $|\zeta| < 1$, and*

$$\begin{aligned}\frac{g^{(n)}(0)}{n!} = \frac{(1 - |z|^2)^n f^{(n)}(z)}{n!} &- C_{n-1,1}\bar{z}\frac{(1 - |z|^2)^{n-1} f^{(n-1)}(z)}{(n-1)!} \\ &+ C_{n-1,2}\bar{z}^2\frac{(1 - |z|^2)^{n-2} f^{(n-2)}(z)}{(n-2)!} - \cdots \\ &+ (-1)^{n-1}\bar{z}^{n-1}(1 - |z|^2)f'(z).\end{aligned} \tag{2.3}$$

Let us write equation (2.2) for $n = k$ and allow $k$ to assume the values $1, 2, \cdots, n$:

$$\frac{(1 - |z|^2)^k f^{(k)}(z)}{k!} = \frac{g^{(k)}(0)}{k!} + C_{k-1,1}\bar{z}\frac{g^{(k-1)}(0)}{(k-1)!} + C_{k-1,2}\bar{z}^2\frac{g^{(k-2)}(0)}{(k-2)!} + \cdots + \bar{z}^{k-1}g'(0). \tag{2.4}$$

Let us proceed similarly with (2.3):

$$\begin{aligned}\frac{g^{(k)}(0)}{k!} = \frac{(1 - |z|^2)^k f^{(k)}(z)}{k!} &- C_{k-1,1}\bar{z}\frac{(1 - |z|^2)^{k-1} f^{(k-1)}(z)}{(k-1)!} \\ &+ C_{k-1,2}\bar{z}^2\frac{(1 - |z|^2)^{k-2} f^{(k-2)}(z)}{(k-2)!} - \cdots \\ &+ (-1)^{k-1}\bar{z}^{k-1}(1 - |z|^2)f'(z).\end{aligned} \tag{2.5}$$

The lemma will be proved if it can be shown that (2.5) is obtained from (2.4) by solving the latter system for $g^{(k)}(0)/k!$ $(k = 1, 2, \cdots, n)$. To do that it suffices to prove that the matrix of the coefficients of (2.4),

$$\Delta = \left\| \begin{array}{ccccc} 1 & 0 & 0 & \cdots & 0 \\ \bar{z} & 1 & 0 & \cdots & 0 \\ \bar{z}^2 & 2\bar{z} & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \bar{z}^{n-1} & C_{n-1,n-2}\bar{z}^{n-2} & C_{n-1,n-3}\bar{z}^{n-3} & \cdots & 1 \end{array} \right\|$$

and the matrix of the coefficients of (2.5),

$$\Delta' = \left\| \begin{array}{ccccc} 1 & 0 & 0 & \cdots & 0 \\ -\bar{z} & 1 & 0 & \cdots & 0 \\ \bar{z}^2 & -2\bar{z} & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ (-1)^{n-1}\bar{z}^{n-1} & (-1)^{n-2}C_{n-1,n-2}\bar{z}^{n-2} & (-1)^{n-3}C_{n-1,n-3}\bar{z}^{n-3} & \cdots & 1 \end{array} \right\|$$

are inverse matrices, or that $\Delta \cdot \Delta' = I$, $I$ being the unit matrix. Now, it is immediately evident that the elements in the principal diagonal of the product matrix are 1, while the element in the $k$th row and $l$th column, where $k > l$, is

$$\begin{aligned} \bar{z}^{k-l}[C_{k-1,k-l} - C_{k-1,k-l-1}C_{l,1} + C_{k-1,k-l-2}C_{l+1,2} - \cdots \pm C_{k-1,2}C_{k-3,k-l-2} \\ \mp C_{k-1,1}C_{k-2,k-l-1} \pm C_{k-1,k-l}]. \end{aligned} \tag{2.6}$$

The sum (2.6) may be written as follows:

$$\bar{z}^{k-l}C_{k-1,k-l}[1 - C_{k-l,1} + C_{k-l,2} - C_{k-l,3} + \cdots \pm 1] = C_{k-1,k-l}(1-1)^{k-l}\bar{z}^{k-l}$$

which is zero. The case $k < l$ may be treated similarly. This proves that $\Delta \cdot \Delta'$ is the unit matrix. Thus, Lemma 2 is established.

3. **Littlewood-Macintyre theorem.** We proceed to prove Theorem 1; this method is different from those of Littlewood and Macintyre. Indeed, form the function

$$\phi(\zeta) = \frac{f((\zeta + z)/(1 + \bar{z}\zeta)) - f(z)}{(1 - |z|^2)f'(z)} \tag{3.1}$$

for a fixed value of $z$ in $|z| < 1$(8). This function is evidently regular and univalent in $|\zeta| < 1$ and omits there the value $(\omega - f(z))/(1-|z|^2)f'(z)$. Since, furthermore, $\phi(0) = 0$ and $\phi'(0) = 1$, we may apply a well known result(9) of Koebe in the theory of univalent functions, according to which

$$\left| \frac{\omega - f(z)}{(1 - |z|^2)f'(z)} \right| \geq \frac{1}{4}.$$

(8) The function $\phi(\zeta)$ plays an important role in the theory of univalent functions, cf. P. Montel, *Leçons Sur Les Fonctions Univalentes ou Multivalentes*, Paris, 1933, p. 51.

(9) See, for example, P. Montel, loc. cit., p. 50.

This proves the theorem. Direct computation shows that the limit is attained for the univalent function $z/(1-z)^2$ and $\omega=-1/4$. Of course Koebe's theorem is the special case $z=0$ of Theorem 1.

4. **Inequalities concerning** $D_1$. For the sequel it is desirable to restate Theorem 1 in a more geometric form. If we set $w=f(z)$, the right side of inequality (1.1) attains its least value when $\omega$ is one of those boundary points of the region $R$ onto which $f(z)$ maps the circle $|z|<1$ which are nearest the point $w$. In that case $|\omega-f(z)|=D_1(w)$, as defined in the introduction, and Theorem 1 becomes

THEOREM 1′. *Let $f(z)$ be analytic and univalent in $|z|<1$. Then the inequality*

$$(1-|z|^2)|f'(z)| \leqq 4D_1(w) \tag{4.1}$$

*is satisfied for all values of $z$ in $|z|<1$, where $D_1(w)$ is the radius of univalence at the point $w=f(z)$ of the region $R$ onto which $f(z)$ maps the circle $|z|<1$.*

It may be of some interest to point out a geometric interpretation of the left side of inequality (4.1). Denote by $\rho(w)$ the "inner radius" of $R$ with respect to a fixed interior point $w$(10). Then $\rho(w)$ can be expressed in terms of $f(z)$ as follows

$$\rho(w) = |f'(z)|(1-|z|^2), \tag{4.2}$$

where $z$ is the point corresponding to $w$. Inequality (4.1) may, therefore, be written in the geometric form (11)

$$\rho(w) \leqq 4D_1(w). \tag{4.3}$$

Theorem 1′ gives an upper bound for $|f'(z)|(1-|z|^2)$. It is desirable also to obtain a lower bound for this expression.

THEOREM 2. *Let $f(z)$ be analytic in $|z|<1$, let $z_0$ be any point of $|z|<1$, and $w_0=f(z_0)$. Then*

$$D_1(w_0) \leqq |f'(z_0)|(1-|z_0|^2). \tag{4.4}$$

We notice that unlike (4.1), the relation (4.4) holds without any restriction other than analyticity on the function $f(z)$. Denote by $R$ the Riemann surface over the $w$-plane onto which $w=f(z)$ maps the circle $|z|<1$. If $w_0$ is

(10) The "inner radius" of a simply connected region $R$ with respect to an interior point $w_0$ is the radius of the circle on which the region $R$ can be mapped conformally by a function $f(w)$ so that $f(w_0)=0$ and $f'(w_0)=1$. Cf. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze*, vol. II, Berlin, 1925, pp. 16–21.

(11) Inequalities (4.1) and (4.3) together with Corollary 2 below were first proved by J. L. Walsh, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 520–523. In the same paper the author suggests the use of the present method in the study of higher derivatives of univalent functions, which is one of the principal topics taken up in the present chapter.

a branch point of $R$, (4.4) is trivial, for in that case both sides of the inequality reduce to zero. Otherwise, let

$$g(\zeta) = f\left(\frac{\zeta + z_0}{1 + \bar{z}_0\zeta}\right).$$

This function is also analytic in $|\zeta| < 1$ and maps the circle onto $R$. Furthermore, $g(0) = w_0$. If we denote by $\zeta = h(w)$ the inverse function of $w = g(\zeta)$, the function $h(w)$ is defined, regular, and single-valued on $R$. In particular, a suitable branch of $h(w)$ will be regular and single-valued on the single-sheeted circle $C$ with center at $w_0$ and radius $D_1(w_0)$. The values which this branch assumes in $C$ all lie in the circle $|\zeta| < 1$. Hence, in $C$: $|h(w)| < 1$, $h(w_0) = 0$. Consequently, applying Schwarz's lemma

$$|h'(w_0)| \leqq \frac{1}{D_1(w_0)}.$$

Hence, $|g'(0)| \geqq D_1(w_0)$ and the evaluation of $g'(0)$ in terms of $f(z)$ yields (4.4).

The inequality in (4.4) is sharp, reducing to an equality when

$$f(z) = \frac{z - z_1}{1 - \bar{z}_1 z}, \qquad |z_1| < 1.$$

Combining Theorems 1′ and 2, we obtain

THEOREM 3. *Let $f(z)$ be regular and univalent in $|z| < 1$, let $z_0$ be any point of $|z| < 1$, and $w_0 = f(z_0)$. Then,*

$$D_1(w_0) \leqq |f'(z_0)|(1 - |z_0|^2) \leqq 4D_1(w_0). \tag{4.5}$$

We remark that Theorems 1 and 1′ can be somewhat improved if we assume $f(z)$ not merely analytic and univalent in $|z| < 1$, but also bounded there: $|f(z)| \leqq M$. Under those conditions the function $\phi(z)$ defined by (3.1) is also analytic and univalent there, with $\phi(0) = 0$, $\phi'(0) = 1$,

$$|\phi(\zeta)| \leqq \frac{2M}{|f'(z)|(1 - |z|^2)}.$$

Since $\phi(\zeta)$ in $|\zeta| < 1$ omits the value

$$\frac{\omega - f(z)}{f'(z)(1 - |z|^2)}$$

provided the function $f(z)$ omits the value $\omega$, the inequality of Pick([12]) yields

([12]) That is to say, under a smooth map of the region $|z| < 1$ by a function $w = f(z)$ with $f(0) = 0$, $f'(0) = 1$, $|f(z)| < M$, every boundary point of the image in the $w$-plane satisfies the inequality $|w| \geqq [M - (M^2 - M)^{1/2}]^2$. See Pick, and R. Nevanlinna, loc. cit.

$$D_1(w_0) \geqq \left[\frac{2M}{|f'(z_0)|^{1/2}(1-|z_0|^2)^{1/2}} - \left(\frac{4M^2}{|f'(z_0)|(1-|z_0|^2)} - 2M\right)^{1/2}\right]^2,$$

$$\frac{4M[D_1(w_0)]^{1/2}}{D_1(w_0)+2M} \geqq |f'(z_0)|^{1/2}(1-|z_0|^2)^{1/2}.$$

It may be noted that as $M$ becomes infinite this last inequality approaches the form (4.1).

5. **Applications.** From Theorem 3 various corollaries may be immediately deduced.

COROLLARY 1. *Let $f(z)$ be regular and univalent in $|z|<1$, $\{z_n\}$ any sequence of points in $|z|<1$, and $w_n=f(z_n)$. Then, a necessary and sufficient condition that*

$$\lim_{n\to\infty} |f'(z_n)|(1-|z_n|) = 0$$

*is that*

$$\lim_{n\to\infty} D_1(w_n) = 0,$$

*and a necessary and sufficient condition that $|f'(z_n)|(1-|z_n|)$ remain bounded is that $D_1(w_n)$ remain bounded.*

COROLLARY 2. *Let $f(z)$ be regular, univalent, and bounded in $|z|<1$, $\{z_n\}$ any sequence of points in $|z|<1$ for which $\lim_{n\to\infty}|z_n|=1$. Then*

$$\lim_{n\to\infty} |f'(z_n)|(1-|z_n|)=0.$$

The proof of Corollary 1 follows directly from the inequalities (4.5), while Corollary 2 follows from Corollary 1 if one remarks that under the hypotheses of Corollary 2 we have $D_1(w_n)\to 0$(13). Another consequence of (4.5) is the following:

COROLLARY 3. *Let $f(z)$ be regular and univalent in $|z|<1$, let $z_0$ be any point of $|z|=1$. Then there exists a sequence of points $\{z_n\}$ ($|z_n|<1$) converging to $z_0$ such that*

$$\lim_{n\to\infty} |f'(z_n)|(1-|z_n|) = 0.$$

In accordance with Corollary 1 it suffices to find a sequence $\{z_n\}$ converging to $z_0$ for which the points $w_n=f(z_n)$ satisfy the relation $D_1(w_n)\to 0$. Such a

(13) As was pointed out by Walsh (loc. cit.), Corollary 2 may also be proved by Carathéodory's method of the conformal mapping of variable regions, cf. C. Carathéodory, *Conformal Representation*, Cambridge, 1932, p. 75.

sequence may be found as follows. It is well known[14] that a univalent function has finite limit values on almost all radii. These limit values are boundary points of the region onto which $f(z)$ maps the circle $|z| < 1$. Choose a sequence of such radii $r_n$ which converges to the radius joining $z_0$ with the origin. On the radius $r_n$ choose a point $z_n$ ($|z_n| < 1$) so near to the circumference $|z| = 1$ that

$$D_1(w_n) < 1/n.$$

This sequence $\{z_n\}$ fulfills the necessary requirements.

6. **Inequalities for higher derivatives.** We now turn to the corresponding study of the higher derivatives of univalent functions. In particular, we shall determine upper bounds for expressions of the form

$$|f^{(n)}(z_0)|(1 - |z_0|^2)^n. \tag{6.1}$$

It is clear immediately that lower bounds for these expressions in terms of $D_1(w)$ cannot be obtained even in the case $n=2$. For the expression (6.1) is identically zero for $n \geqq 2$ when $f(z) \equiv z$. Even for the upper bounds of (6.1) the sharp inequalities will now be obtained only in the case $n = 2, 3$. For higher values of $n$ the corresponding inequalities depend on the assumption of the truth of Bieberbach's conjecture, which up to the present has not been established.

We begin by proving the following inequalities

THEOREM 4. *Let $f(z)$ be regular and univalent in $|z| < 1$, let $z_0$ be any point of $|z| < 1$, and let $w_0 = f(z_0)$. Then,*

$$|f''(z_0)|(1 - |z_0|^2)^2 \leqq 8(|z_0| + 2)D_1(w_0) \tag{6.2}$$

*and*

$$|f'''(z_0)|(1 - |z_0|^2)^3 \leqq 24(|z_0|^2 + 4|z_0| + 3)D_1(w_0). \tag{6.3}$$

*These inequalities are sharp, reducing to equalities for $f(z) = z/(1+z)^2$ for real negative values of $z$.*

To prove (6.2) and (6.3) compute the second and third Taylor coefficients, $b_2$ and $b_3$, of the function (3.1) where we set $z = z_0$. By direct computation (or by §2, Lemma 2) we find that

$$\begin{aligned} b_2 &= \frac{1}{2}\frac{f''(z_0)}{f'(z_0)}(1 - |z_0|^2) - \bar{z}_0, \\ b_3 &= \frac{1}{6}\frac{f'''(z_0)}{f'(z_0)}(1 - |z_0|^2)^2 - \frac{f''(z_0)}{f'(z_0)}\bar{z}_0(1 - |z_0|^2) + \bar{z}_0^2. \end{aligned} \tag{6.4}$$

---

[14] See, for example, W. Seidel, Mathematische Annalen, vol. 104 (1931), p. 191.

Now, according to Bieberbach's theorem and Löwner's theorem[15] $|b_2| \leqq 2$ and $|b_3| \leqq 3$. Hence

$$\left| \frac{1}{2} \frac{f''(z_0)}{f'(z_0)} (1 - |z_0|^2) - \bar{z}_0 \right| \leqq 2$$

and

$$(6.2') \qquad |f''(z_0)(1 - |z_0|^2)^2 - 2\bar{z}_0(1 - |z_0|^2)f'(z_0)| \leqq 4(1 - |z_0|^2)\,|f'(z_0)|.$$

Applying (4.1) we obtain at once inequality (6.2). To obtain (6.3) we use the evaluation of $b_3$ in (6.4) and write

$$\left| \frac{1}{6} \frac{f'''(z_0)}{f'(z_0)} (1 - |z_0|^2)^2 - \frac{f''(z_0)}{f'(z_0)} \bar{z}_0(1 - |z_0|^2) + \bar{z}_0^2 \right| \leqq 3$$

and

$$|f'''(z_0)(1 - |z_0|^2)^3 - 6\bar{z}_0 f''(z_0)(1 - |z_0|^2)^2 + 6\bar{z}_0^2 f'(z_0)(1 - |z_0|^2)|$$
$$\leqq 18\,|f'(z_0)|\,(1 - |z_0|^2).$$

It follows that

$$\begin{aligned} |f'''(z_0)|\,(1 - |z_0|^2)^3 &\leqq |6\bar{z}_0 f''(z_0)(1 - |z_0|^2)^2 - 6\bar{z}_0^2 f'(z_0)(1 - |z_0|^2)| \\ &\quad + 18\,|f'(z_0)|\,(1 - |z_0|^2) \\ &\leqq 6\,|\bar{z}_0 f''(z_0)(1 - |z_0|^2)^2 - 2\bar{z}_0^2 f'(z_0)(1 - |z_0|^2)| \\ &\quad + 6\,|z_0|^2\,|f'(z_0)|\,(1 - |z_0|^2) + 18\,|f'(z_0)|\,(1 - |z_0|^2). \end{aligned}$$

Applying now inequalities (6.2′) and (4.1) we obtain inequality (6.3).

If now Bieberbach's conjecture concerning the coefficients of univalent were known to be true[16], one could write

$$\frac{|\phi^{(n)}(0)|}{n!} \leqq n.$$

With the aid of a little algebraic manipulation (see below) this would lead to the sharp inequality

$$(6.5) \qquad |f^{(n)}(z_0)|\,(1 - |z_0|^2)^n \leqq 4n!(n + |z_0|)(1 + |z_0|)^{n-2} D_1(w_0),$$

which becomes an equality for $f(z) = z/(1+z)^2$ for real negative values of $z$. Unfortunately, however, the inequality $|b_n| \leqq n$ has been proved only for

[15] L. Bieberbach, Sitzungsberichte der Königlichen Preussischen Akademie der Wissenschaften zu Berlin, vol. 38 (1916), pp. 940–955; K. Löwner, Mathematische Annalen, vol. 89 (1923), pp. 103–121.

[16] See, for example, L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, 2d edition, 1931, p. 80, Footnote 4.

$n=2$ and 3, so that the validity of inequality (6.5) has been established for $n=2$ and 3 only. Weaker inequalities have actually been proved by various authors, in particular, J. E. Littlewood[17] who showed

$$\frac{|\phi^{(n)}(0)|}{n!} < en \tag{6.6}$$

and E. Landau[18] who showed

$$\frac{|\phi^{(n)}(0)|}{n!} \leqq \left(\frac{1}{2} + \frac{1}{\pi}\right) en.$$

Making use of (6.6) and Lemma 1 of §2 we find

$$\frac{(1 - |z|^2)^n |f^{(n)}(z)|}{n!} \leqq (1 - |z|^2) |f'(z)| \sum_{\nu=0}^{n-1} C_{n-1,\nu} |z|^\nu \frac{|\phi^{(n-\nu)}(0)|}{(n-\nu)!}$$

and using (4.1)

$$\frac{(1 - |z|^2)^n |f^{(n)}(z)|}{n!} \leqq 4eD_1(w) \sum_{\nu=0}^{n-1} (n-\nu) C_{n-1,\nu} |z|^\nu.$$

Since, however,

$$\sum_{\nu=0}^{n-1} C_{n-1,\nu} |z|^\nu = (1 + |z|)^{n-1}$$

and

$$\sum_{\nu=0}^{n-1} \nu C_{n-1,\nu} |z|^\nu = (n-1) |z| (1 + |z|)^{n-2},$$

we obtain

$$(1 - |z|^2)^n |f^{(n)}(z)| \leqq 4e \cdot n! D_1(w) [n(1 + |z|)^{n-1} + (n-1) |z| (1 + |z|)^{n-2}]$$

and finally

$$(1 - |z|^2)^n |f^{(n)}(z)| \leqq 4e \cdot n! (n + |z_0|)(1 + |z_0|)^{n-2} D_1(w_0).$$

This clearly is not a sharp inequality. We thus obtain

THEOREM 5. *Let $f(z)$ be regular and univalent in $|z| < 1$, let $z_0$ be any point in $|z| < 1$, and let $w_0 = f(z_0)$. Then*

$$|f^{(n)}(z_0)| (1 - |z_0|^2)^n \leqq 4e \cdot n! (|z_0| + n)(1 + |z_0|)^{n-2} D_1(w_0). \tag{6.7}$$

From this inequality we obtain again two corollaries analogous to those of Theorem 3.

---

(17) J. E. Littlewood, loc. cit., p. 498.

(18) E. Landau, Mathematische Zeitschrift, vol. 30 (1929), p. 635.

COROLLARY 4. *Let $f(z)$ be regular and univalent in $|z|<1$, $\{z_n\}$ any sequence of points in $|z|<1$ and $w_n=f(z_n)$. Then, if*

$$\lim_{n\to\infty} D_1(w_n) = 0,$$

*all the derivatives of $f(z)$ will satisfy the relation*

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \qquad k = 1, 2, 3, \cdots.$$

Clearly the converse of the theorem is false since taking $f(z)\equiv z$, $z_n=0$ $(n=1, 2, \cdots)$, we have $f^{(k)}(z_n)=0$ for all $k\geqq 2$ and all $n$ while $D_1(w_n)=1$.

COROLLARY 5. *Let $f(z)$ be regular, univalent, and bounded in $|z|<1$, $\{z_n\}$ any sequence of points in $|z|<1$ for which $\lim_{n\to\infty}|z_n|=1$. Then*

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \qquad k = 1, 2, 3, \cdots.$$

7. **Applications.** A few remarks concerning Theorem 3 will now be made. Koebe's "Verzerrungssatz" can be written in the form[19]

$$\frac{1-|z|}{(1+|z|)^3} \leqq |f'(z)| \leqq \frac{1+|z|}{(1-|z|)^3}.$$

If we combine this inequality with (4.5) we obtain

$$\frac{1}{4}\left(\frac{1-|z_0|}{1+|z_0|}\right)^2 \leqq D_1(w_0) \leqq \left(\frac{1+|z_0|}{1-|z_0|}\right)^2.$$

We may state this result as follows:

COROLLARY 6. *Let $f(z)$ be regular and univalent in $|z|<1$ with $f(0)=0$, $f'(0)=1$, let $z_0$ be any point of $|z|<1$, and let $w_0=f(z_0)$. Then the radius of univalence $D_1(w_0)$ at the point $w_0$ satisfies the inequality*

$$\frac{1}{4}\left(\frac{1-|z_0|}{1+|z_0|}\right)^2 \leqq D_1(w_0) \leqq \left(\frac{1+|z_0|}{1-|z_0|}\right)^2.$$

The lower bound of $D_1(w_0)$ was obtained in less precise form by W. E. Sewell[20]. The first inequality is sharp, becoming an equality for $f(z) = z/(1+z)^2$ along the positive real axis. The second inequality is probably not sharp.

Another application of Theorem 3 concerns infinite regions. Suppose that $R$ is a simply connected region of the $w$-plane for which $w=\infty$ is an accessible boundary point, let

(19) See, for instance, Paul Montel, loc. cit., p. 52.

(20) W. E. Sewell, these Transactions, vol. 41 (1937), p. 90.

$$\limsup_{w\to\infty} D_1(w) = D,$$

where $w$ is an interior point of $R$, and let $w=f(z)$ map $R$ on the interior of the circle $|z|<1$; suppose that $z=\alpha$, $(|\alpha|=1)$, corresponds to $w=\infty$. From Theorem 3 it follows that

$$\limsup_{z\to\alpha} |f'(z)|(1-|z|^2) \leqq 4D.$$

For an arbitrary infinite region the relations $w_n=f(z_n)\to\infty$, $\limsup_{n\to\infty} D_1(w_n)=D$, $\liminf_{n\to\infty} D_1(w_n)=d$ clearly imply that $\limsup_{n\to\infty}|f'(z_n)|(1-|z_n|^2)\leqq 4D$, $\liminf_{z\to\alpha}|f'(z_n)|(1-|z|^2)\geqq d$.

The final remark concerns an inequality derived by G. Szegö[21] on the difference quotient of a univalent function. His inequality is as follows: Let $f(z)$ be regular and univalent in $|z|<1$, let $z_1$ and $z_2$ be any two points of the circle $|z|<1$. Then,

$$\begin{aligned}(7.1)\qquad &|f'(z_2)|(1-|z_2|^2)\frac{|1-\bar{z}_2z_1|}{(|z_1-z_2|+|1-\bar{z}_2z_1|)^2}\\ &\leqq \left|\frac{f(z_1)-f(z_2)}{z_1-z_2}\right| \leqq |f'(z_2)|(1-|z_2|^2)\frac{|1-\bar{z}_2z_1|}{(|z_1-z_2|-|1-\bar{z}_2z_1|)^2}.\end{aligned}$$

Let us introduce the non-euclidean distance $\rho(z_1, z_2)$ between the points $z_1$ and $z_2$ by means of the following relations

$$\rho(z_1, z_2) = \log\frac{1+r}{1-r}, \qquad r = \left|\frac{z_1-z_2}{1-\bar{z}_2z_1}\right|.$$

By virtue of (7.1) and (4.5) we obtain the inequalities

$$\begin{aligned}D_1(w_2)\frac{|1-\bar{z}_2z_1|}{(|z_1-z_2|+|1-\bar{z}_2z_1|)^2} &\leqq \left|\frac{f(z_1)-f(z_2)}{z_1-z_2}\right|\\ &\leqq 4D_1(w_2)\frac{|1-\bar{z}_2z_1|}{(|z_1-z_2|-|1-\bar{z}_2z_1|)^2},\end{aligned}$$

where $w_2=f(z_2)$. In terms of $\rho(z_1, z_2)$ the inequalities become

$$(7.2)\qquad (1/4)D_1(w_2)(1-e^{-2\rho(z_1,z_2)}) \leqq |f(z_1)-f(z_2)| \leqq (e^{2\rho(z_1,z_2)}-1)D_1(w_2).$$

From the inequalities (7.2) we obtain the corollary:

COROLLARY 7. *Let $f(z)$ be regular and univalent in $|z|<1$, let $\{z_n\}$ and $\{z_n'\}$ be two sequences of points in $|z|<1$, such that $\rho(z_n, z_n')$ is bounded and let $w_n'=f(z_n')$. Then $\lim_{n\to\infty}|f(z_n)-f(z_n')|=0$ if, and only if,*

$$\lim_{n\to\infty}(e^{2\rho(z_n,z_n')}-1)D_1(w_n') = 0.$$

---

[21] G. Szegö, Mathematische Annalen, vol. 100 (1928), pp. 190–191.

8. **Behavior of the first derivative almost everywhere.** Corollary 2 may be stated as asserting that for a regular, univalent, and bounded function in the circle $|z|<1$ the first derivative is of order $o((1-r)^{-1})$ on all radii of the circle. The next theorem shows, however, that this order of growth can be attained only on a small number of radii and that on most radii the order of growth is considerably smaller. Indeed, we prove the following

THEOREM 6. *Let $f(z)$ be regular and univalent in the circle $|z|<1$. Then*

$$(8.1)\qquad \lim_{z\to e^{i\alpha}} |f'(z)|(1-|z|)^{1/2} = 0$$

*for all points $e^{i\alpha}$ of the circumference $|z|=1$ with the exception of at most a set of measure zero, where $z$ in the above limit is taken in any angle less than $\pi$ with vertex in $e^{i\alpha}$ and bisected by the radius joining $z=0$ with $z=e^{i\alpha}$. Furthermore, in any such angle the above limit is uniform.*

The proof depends on a number of lemmas.

LEMMA 3. *If $f(z)$ is univalent in the circle $|z|<1$, then on almost all radii*

$$(8.2)\qquad |f'(z)| = O((1-|z|)^{-1/2}),$$

*where the symbol $O$ does not necessarily indicate uniformity for the different radii. The relation* (8.2) *holds also in any angle of the type described in Theorem* 6 *which corresponds to a radius for which* (8.2) holds.

If we set $w=f(z)$, then the function maps $|z|<1$ on a simply connected region $R$ of the $w$-plane. Now, this region $R$ possesses at least two distinct boundary points $w=a$ and $w=b$, $(a\neq b)$. Indeed, if $R$ were the entire plane then the inverse function $z=g(w)$ of $w=f(z)$ would map the plane on the interior of $|z|<1$. It would, therefore, be bounded in the whole plane and by Liouville's theorem be identically a constant, which is contrary to our assumption. If $R$ were the whole plane with the exception of one point, $w=a$, then $g(w)$ would be regular and bounded in the whole plane with the exception of the one point, $w=a$. This point, by Riemann's theorem, would be a removable singularity, and again $z=g(w)$ would be identically constant. Now, by a familiar argument the function

$$t = \frac{1}{((w-a)/(w-b))^{1/2}-c} = \lambda(w),$$

where the constant $c$ is suitably chosen, maps the region $R$ conformally on a bounded region of the $t$-plane.

The function

$$h(z) = \lambda(f(z))$$

is regular, univalent and bounded in $|z|<1$. Let us suppose that Lemma 3

has already been proved for $h(z)$. Then, it will also hold for $f(z)$. Indeed,

$$f'(z) = \frac{h'(z)}{\lambda'(f(z))}.$$

Since we have assumed that $\limsup_{z\to e^{i\alpha}} |h'(z)| (1-|z|)^{1/2} < \infty$ for almost all points $z=e^{i\alpha}$ on $|z|=1$, where $z$ lies in corresponding angles as described in Theorem 6, the asserted lemma will follow for $f'(z)$ provided that $\liminf_{z\to e^{i\alpha}} |\lambda'(f(z))| > 0$ for almost all $e^{i\alpha}$ in the corresponding angles. But now

$$\lambda'(w) = -\frac{a-b}{2}\frac{[\lambda(w)]^2}{(w-b)^{3/2}(w-a)^{1/2}},$$

which shows that $\liminf_{z\to e^{i\alpha}} |\lambda'(f(z))| = 0$ only if there exists a sequence of points $z_n \to e^{i\alpha}$ for which $f(z_n)\to b$ or $f(z_n)\to a$. This, however, can only happen for a set of $e^{i\alpha}$ of measure zero(22).

It suffices, therefore, to prove Lemma 3 for a bounded univalent function $f(z)$. Now, $w=f(z)$ maps the circle $|z|<1$ on a bounded region of the $w$-plane. Denote the area of this region by $A$. We have, setting $z=re^{i\theta}$,

$$(8.3) \qquad \int_0^{2\pi}\int_0^{\rho} |f'(re^{i\theta})|^2 r\, dr\, d\theta < A$$

for every $0 \leqq \rho < 1$. The function

$$(8.4) \qquad \Phi(z) = \int_0^z z[f'(z)]^2 dz$$

is regular in $|z|<1$ and we shall perform the integration along the radius joining $z=0$ and $z=re^{i\theta}$ so that

$$\Phi(\rho e^{i\theta}) = \int_0^{\rho} re^{2i\theta}[f'(re^{i\theta})]^2 dr.$$

Hence,

$$|\Phi(\rho e^{i\theta})| \leqq \int_0^{\rho} r|f'(re^{i\theta})|^2 dr.$$

Integration of the last inequality with respect to $\theta$ together with (8.3) yields

$$(8.5) \qquad \int_0^{2\pi} |\Phi(\rho e^{i\theta})|\, d\theta < A$$

for every $0 \leqq \rho < 1$.

Now it is a familiar fact that if a function $\Phi(z)$ is regular in $|z|<1$ and

(22) F. and M. Riesz, Compte Rendu du Quatrième Congrès des Mathématiciens Scandinaves, 1920, pp. 28–30.

satisfies the condition (8.5) it may be represented in the following form([23]):

$$\Phi(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{it}+z}{e^{it}-z}\, d\mu(t) + i\beta, \tag{8.6}$$

where the integral is a Stieltjes integral, $\mu(t)$ is a function of bounded variation in the interval $0 \leqq t \leqq 2\pi$ and $\beta$ is a constant. Equations (8.4) and (8.6) permit us to express $[f'(z)]^2$ in the form

$$[f'(z)]^2 = \frac{1}{\pi z}\int_0^{2\pi} \frac{e^{it}}{(e^{it}-z)^2}\, d\mu(t).$$

Hence,

$$(1-r^2)\,|f'(z)|^2 \leqq \frac{1}{\pi r}\int_0^{2\pi} \frac{1-r^2}{1-2r\cos(\theta-t)+r^2}\, dM(t), \tag{8.7}$$

where $M(t)$ denotes the total variation of the function $\mu(t)$ in the interval $(0, t)$. The right-hand side of this inequality approaches a definite finite limit as $z = re^{i\theta} \to e^{i\alpha}$ in an angle of the type described in Theorem 6 for almost all $e^{i\alpha}$([24]). Hence, the right-hand side remains bounded in such angles. Thus,

$$|f'(z)| \leqq C_\alpha \cdot (1 - |z|)^{-1/2} \tag{8.8}$$

in the angular neighborhood of almost all points $e^{i\alpha}$, where $C_\alpha$ is a constant independent of $z$, but in general depending on $\alpha$. This proves the lemma.

COROLLARY 8. *Let $w = f(z)$ be regular and univalent in $|z| < 1$. Then, for almost all points $e^{i\alpha}$ on $|z| = 1$ every line segment joining an interior point of $|z| < 1$ with $e^{i\alpha}$ is mapped on a rectifiable arc by the function $w = f(z)$.*

This follows readily by integrating (8.8) along such a line segment([25]).

If we restrict ourselves to radial approach in Corollary 8, it is possible to state a sharper result which will be used in the proof of Theorem 6:

LEMMA 4. *Let $w = f(z)$ be regular and univalent in $|z| < 1$. If*

$$l_{\rho,\theta} = \int_\rho^1 |f'(re^{i\theta})|\, dr, \qquad z = re^{i\theta}, \tag{8.9}$$

*then for almost all values of $\theta$ in $0 \leqq \theta \leqq 2\pi$ and for all values of $\rho$ in $0 \leqq \rho < 1$, $l_{\rho,\theta}$ is finite and*

$$\lim_{\rho \to 1} l_{\rho,\theta}(1-\rho)^{-1/2} = 0. \tag{8.10}$$

---

([23]) See, for example, R. Nevanlinna, *Eindeutige analytische Funktionen*, Berlin, 1936, p. 185.

([24]) After integration by parts the integral in (8.7) becomes one of the type considered in Carathéodory's proof of Fatou's theorem. Cf. L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, 2d edition, (1931), pp. 148–151.

([25]) The corollary includes a result stated by M. Lavrentieff, Physico-Mathematical Institute of Stekloff, vol. 5 (1934), p. 207.

The formula in (8.9) represents the length of the image of the radial segment joining the points $\rho e^{i\theta}$ and $e^{i\theta}$.

One may assume without loss of generality, for the same reasons as in the proof of Lemma 3, that $f(z)$ is bounded in $|z|<1$. Then, inequality (8.3) holds for some $A$. The total area $A$ of the image of $|z|<1$ is given by

$$A = \int_0^{2\pi} \int_0^1 |f'(re^{i\theta})|^2 r\, dr\, d\theta.$$

Hence, by Fubini's theorem, for almost all $\theta$ in $0 \leqq \theta \leqq 2\pi$

$$\int_0^1 r\,|f'(re^{i\theta})|^2 dr$$

has a finite value. Hence, for almost all $\theta$

$$\lim_{\rho\to 1} \int_\rho^1 r\,|f'(re^{i\theta})|^2 dr = 0.$$

Thus to any $\epsilon>0$ one may assign a number $\delta=\delta(\epsilon, \theta)$ so that $1-\rho<\delta$ implies

$$\int_\rho^1 r\,|f'(re^{i\theta})|^2 dr < \epsilon$$

for almost all $\theta$. Hence, by Schwarz's inequality

$$\int_\rho^1 |f'(re^{i\theta})|\, dr \leqq \left[(1-\rho)\int_\rho^1 |f'(re^{i\theta})|^2 dr\right]^{1/2} < \left[\frac{\epsilon(1-\rho)}{\rho}\right]^{1/2}$$

for almost all $\theta$ and $1>\rho>1-\delta(\epsilon, \theta)$. This proves (8.10).

Using this lemma, one can now prove

LEMMA 5. *Let $w=f(z)$ be regular and univalent in $|z|<1$. Then on almost all radii*

$$|f'(z)| = o((1-r)^{-1/2}), \qquad |z| = r,$$

*where the symbol o is not intended to indicate uniformity for the different radii.*

We know that on almost all radii (8.2) and (8.10) hold and $\lim_{r\to 1} f(re^{i\theta})=\omega$ exists and is finite([28]). Choose any one of these radii $\theta=\theta_0$ and on it an arbitrary point $z_0$. Let $f(z_0)=w_0$. The segment of the radius between the points $z_0$

([28]) For the proof of the last statement one need merely apply the fact that the integral in (8.9) remains finite for almost all $\theta$. Indeed, take any such $\theta_0$. Then,

$$|f(r_1e^{i\theta_0}) - f(r_2e^{i\theta_0})| \leqq \int_{r_1}^{r_2} |f'(re^{i\theta_0})|\, dr, \qquad r_1 < r_2,$$

and the last integral may be made smaller than any preassigned $\epsilon>0$ provided that $r_1$ and $r_2$ are both chosen sufficiently near unity.

and $e^{i\theta_0}$ is carried into a rectifiable arc joining the points $w_0$ and $\omega$. Its length $l_{z_0}$ is given by

$$l_{z_0} = \epsilon_{z_0}(1 - |z_0|)^{1/2}, \tag{8.11}$$

where by (8.10)

$$\lim_{z_0 \to e^{i\theta_0}} \epsilon_{z_0} = 0, \tag{8.12}$$

the approach being taken radially. Now draw a circle $K_{z_0}$ about the point $z_0$ as center with radius equal to $1-|z_0|$. The interior of the circle $K_{z_0}$ is carried by $w=f(z)$ into a region $R_{z_0}$ of the $w$-plane.

According to Koebe's "Verzerrungssatz" the region $R_{z_0}$ contains the circle $|w-w_0| < ((1-|z_0|)/4)\,|f'(z_0)|$.

Now if we set

$$|f'(z_0)| = C_{z_0}(1 - |z_0|)^{-1/2},$$

according to (8.2) $C_{z_0}$ is bounded along the radius $\theta=\theta_0$. Thus $R_{z_0}$ contains the circle $|w-w_0| < (1/4)C_{z_0}\,(1-|z_0|)^{1/2}$. In view of (8.11) this may also be written $|w-w_0| < C_{z_0}l_{z_0}/4\epsilon_{z_0}$. Denoting by $\rho_{z_0}$ the radius of this circle, we have on the one hand

$$\rho_{z_0} = \frac{C_{z_0}}{4\epsilon_{z_0}}\, l_{z_0}$$

and on the other $\rho_{z_0} \leqq l_{z_0}$. Hence,

$$C_{z_0} \leqq 4\epsilon_{z_0}.$$

Together with (8.12) this implies that

$$\lim_{z_0 \to e^{i\theta_0}} C_{z_0} = 0$$

with radial approach. This proves the lemma.

We are now ready for the proof of Theorem 6. Let $\theta=\theta_0$ be a radius for which (8.2) holds in any angle as asserted in Lemma 3 and also

$$\lim_{r \to 1} |f'(re^{i\theta_0})|\,(1 - r)^{1/2} = 0. \tag{8.13}$$

By Lemmas 3 and 5 the set of such $\theta_0$ is of measure $2\pi$.

Consider the function

$$g(z) = f'(z)(e^{i\theta_0} - z)^{1/2},$$

where we choose that branch of the square root which is positive for real positive values of the radicand. This function is regular and single-valued in $|z| < 1$. Now, take a fixed angle of opening less than $\pi$ with vertex in $e^{i\theta_0}$. In this angle

$$\frac{1}{M} < \frac{|e^{i\theta_0} - z|}{1 - |z|} < M$$

for a suitable positive constant $M$. Hence, by (8.13)

$$\lim_{r \to 1} g(re^{i\theta_0}) = 0,$$

while by (8.2) the function $g(z)$ is bounded in the fixed angle. By Lindelöf's theorem ([27]) $\lim_{z \to e^{i\theta_0}} g(z) = 0$ uniformly in every angle contained in the fixed angle. This proves the theorem.

9. **Example on the slowness of approach of $|f^{(k)}(z)|\ (1-|z|)^k$.** We have shown in Corollary 2, §5, that if the function $f(z)$ is bounded and univalent for $|z| < 1$, and also under various alternative conditions, then we have

$$\lim_{|z_n| \to 1} f'(z_n)(1 - |z_n|) = 0, \qquad |z_n| < 1. \tag{9.1}$$

Even for the class of bounded univalent functions, continuous in $|z| \leqq 1$, equation (9.1) cannot be improved by establishing results on rate of approach in equation (9.1) or by replacing the second factor by that factor raised to a suitable power. Indeed we shall prove that the limit in (9.1) can be approached arbitrarily slowly, in the sense of

THEOREM 7. *Let the function $Q(r)$ be defined and positive for $0 < r < 1$, with $\lim_{r \to 1} Q(r) = 0$. Then there exists a function $F(z)$ analytic and univalent interior to $\gamma$: $|z| = 1$, continuous for $|z| \leqq 1$, and there exists a sequence of points $z_1, z_2, \cdots$ interior to $\gamma$ with $|z_n| = r_n \to 1$, such that we have*

$$\lim_{n \to \infty} \frac{F'(z_n)(1 - |z_n|)}{Q(|z_n|)} = \infty. \tag{9.2}$$

*In fact, we shall choose $F(z)$ real for real $z$, and $z_n$ real.*

As a matter of convenience, we establish first Theorem 7 and then an extension of Theorem 7 to higher derivatives. The ensuing proof is given in preparation for the more general theorem, and is somewhat more complicated than is necessary for the proof of Theorem 7 alone.

We shall find useful a function analytic and univalent for $|z| < 1$ whose Taylor expansion about the origin has all of its coefficients positive. Such a function is

$$w_1 = f_1(z) = \frac{z}{(1 - z)^2} = z + 2z^2 + 3z^3 + \cdots$$

which maps the region $|z| < 1$ smoothly onto the $w_1$-plane slit along the axis of reals from $-1/4$ to $-\infty$. The function

([27]) E. Lindelöf, Acta Societatis Scientiarum Fennicae, vol. 46 (1915).

$$w_2 = f_2(z) = \frac{f_1(\rho z)}{\rho} = z + 2\rho z^2 + 3\rho^2 z^3 + \cdots, \qquad 0 < \rho < 1,$$

then maps $|z| < 1$ smoothly onto a *Jordan region*([28]) symmetric in the axis of reals. For definiteness we choose $\rho = 1/2$, and denote by $J_0$ the Jordan region of the $w$-plane which is the image of $|z| < 1$ under the map([29])

$$w = F_0(z) = z + \frac{2}{2} z^2 + \frac{3}{2^2} z^3 + \frac{4}{2^3} z^4 + \cdots. \tag{9.3}$$

Construct in the $w$-plane new Jordan regions $J_1, J_2, \cdots$ with the same shape and orientation as $J_0$, mutually exterior and exterior to $J_0$, with the analogue $B_k$ for $J_k$ of the point $w = 0$ for $J_0$ lying on the axis of reals, so that the sequence $B_0 = 0, B_1, B_2, \cdots$ forms a monotonically increasing sequence. Choose moreover the region $J_k$ just $(1/2^k)$th the size of $J_0$ in linear dimensions, and locate (as is possible) the sequence of regions $J_k$ in such a way that their totality lies in some circle $|w| \leqq D$.

The region $J_0$ is symmetric in the axis of reals, so its boundary (an analytic Jordan curve) cuts that axis in precisely two points $A_0$ (to the left of the origin) and $C_0$ (to the right of the origin). Denote the analogous points for $J_k$ by $A_k$ and $C_k$. The boundary of $J_k$ has a vertical tangent at both $A_k$ and $C_k$.

A Jordan region $R$ is to be constructed in the $w$-plane from the regions $J_0, J_1, J_2, \cdots$ by connecting each region to the preceding region by a canal; each of the two banks of such a canal shall be a segment of one of the lines $y = \pm d_k$, $d_k > 0$. Each point interior to $J_k$ shall lie interior to $R$. The first canal, whose boundaries are segments of $y = \pm d_1$, joins $J_0$ in the neighborhood of $C_0$ with $J_1$ in the neighborhood of $A_1$; the second canal, whose boundaries are segments of $y = \pm d_2$, joins $J_1$ in the neighborhood of $C_1$ with $J_2$ in the neighborhood of $A_2$, and so on. The choice of the numbers $d_k$ is now to be made more precise.

Denote by $w = F(z)$ the function which maps $|z| < 1$ onto $R$ with $F(0) = 0$, $F'(0) > 0$; of course $F(z)$ depends on the numbers $d_1, d_2, \cdots$. Choose $d_1$ independently of $d_2, d_3, \cdots$ so small that the subset $R_1$ composed of all points of $R$ not in $J_0$ corresponds under the transformation $w = F(z)$ to a set of points $z$ interior to $\gamma$: $|z| = 1$ at which we have

$$Q(|z|) < 1/3. \tag{9.4}$$

([28]) A Jordan region is any region bounded by a Jordan curve.

([29]) It is sufficient for the purpose of both Theorem 7 and Theorem 8 to choose here a function $F_0(z)$ which maps $|z| < 1$ smoothly onto a Jordan region with $F_0(0) = 0$, $F_0'(0) = 1$, and has all of the coefficients of its Taylor expansion about the origin positive. For instance we may also choose

$$w = F_0(z) = \frac{2z}{2 - z} = z + \frac{1}{2} z^2 + \frac{1}{2^2} z^3 + \cdots$$

which maps $|z| < 1$ onto the interior of the circle $|w - 2/3| = 4/3$.

Such choice of $d_1$ is possible. For under the map $w=F(z)$ it follows from a theorem due to Lindelöf([30]) that the subset $R_1$ is mapped into a set bounded in part by an arc of $\gamma$ and whose remaining boundary (a Jordan arc) can be made as near to $\gamma$ as desired. For the boundary points of $R_1$ not boundary points of $R$ are the points of the boundary of $J_0$ in the neighborhood of the point $C_0$ between the lines $y=\pm d_1$; by choosing $d_1$ sufficiently small all such points can be made uniformly as near as desired to the boundary of $R$; so by Lindelöf's theorem all points of the boundary of the transform of $R_1$ (and hence all points of the transform of $R_1$ itself) can be made as near to $\gamma$ as desired, and (9.4) is justified.

Similarly the number $d_2$ is to be chosen so small that all points of $R$ not in $J_0$ or $J_1$ or in the canal joining $J_0$ and $J_1$ correspond under the map $w=F(z)$ to points interior to $\gamma$ at which we have $Q(|z|)<1/9$; more generally the number $d_k$ is to be chosen so that all points of $R$ not in $J_0 \cdot J_1, \cdots, J_{k-1}$ or in the canals joining successive regions $J_0, J_1, \cdots, J_{k-1}$, correspond under the map $w=F(z)$ to points interior to $\gamma$ at which we have

$$Q(|z|) < 1/3^k; \tag{9.5}$$

such successive choice of the numbers $d_k$ is possible, again by Lindelöf's theorem. There are no further restrictions on the numbers $d_k$ so far as the requirements of Theorem 7 itself are concerned. We now introduce the inner radius $\rho(w_0)$ of the region $R$ with respect to the arbitrary point $w_0$ of $R$([31]). It is well known that $\rho(w_0)$ has a monotonic character with respect to $R$: if $R$ is increased so also is $\rho(w_0)$; if $R$ is stretched uniformly in the linear ratio $1:m$ with $w_0$ fixed, then $\rho(w_0)$ is multiplied by $m$; if $R$ is the interior of a circle with center at $w_0$, the inner radius is the usual radius of this circle.

The inner radius of $R$ with respect to the point $B_k$ is greater than $1/2^k$, for it follows from (9.3) that the inner radius of $J_0$ with respect to $B_0$ is unity, so the inner radius of $J_k$ with respect to $B_k$ is $1/2^k$. On the other hand, if $z_k$ denotes the point of $|z|<1$ which corresponds to the point $B_k$ under the transformation $w=F(z)$, the inner radius of $R$ with respect to $B_k$ is $|F'(z_k)|(1-|z_k|^2)$, so we may write $|F'(z_k)|(1-|z_k)^2)>1/2^k$. From inequality (9.5) we have $Q(|z_k|)<1/3^k$, whence

$$\frac{|F'(z_k)|(1-|z_k|^2)}{Q(|z_k|)} > \frac{3^k}{2^k}, \tag{9.6}$$

from which (9.2) follows([32]).

([30]) Acta Societatis Scientiarum Fennicae, vol. 46 (1915). Or see Walsh, *Interpolation and Approximation*, §2.1. In applying Lindelöf's result it is essential to notice that the region $R$ is bounded independently of the numbers $d_k$.

([31]) Cf. §4, Footnote 10.

([32]) In the proof of Theorem 7 we might equally well have used an example due to Szegö, Mathematische Zeitschrift, vol. 23 (1925), pp. 45–61; pp. 57–59. Szegö does not mention the

Under the present circumstances the region $R$ is symmetric in the axis of reals, the numbers $z_k$ are real, and $F'(z_k)$ is positive, so the absolute value signs may be removed from (9.6). Of course $F(z)$ is continuous in $|z| \leq 1$ (when suitably defined on $|z| = 1$), as the mapping function for a Jordan region. The points $B_k$ are real and positive and approach the boundary of $R$, so the points $z_k$ are real and positive and approach the point $z = 1$.

Theorem 7 shows that the limit in (9.1) can be approached arbitrarily slowly; by virtue of §4, Theorem 3, we may also say that $\lim_{|z_n| \to 1} D_1[f(z_n)]$ considered as a function of $1 - |z_n|$ can also be approached arbitrarily slowly.

We now consider the generalization of Theorem 7 to higher derivatives:

THEOREM 8. *Let the function $Q(r)$ be defined and positive for $0 < r < 1$, with $\lim_{r \to 1} Q(r) = 0$. Let the positive integer $m$ be given. Then there exists a function $F(z)$ analytic and univalent interior to $\gamma$: $|z| = 1$, continuous for $|z| \leq 1$, and a sequence of points $z_1, z_2, \cdots$ interior to $\gamma$ with $|z_n| = r_n \to 1$, such that we have*

$$\lim_{n \to \infty} \frac{F^{(m)}(z_n)(1 - |z_n|)^m}{Q(|z_n|)} = \infty. \tag{9.7}$$

*Indeed, we shall choose $F(z)$ real for real $z$, and $z_n$ real.*

In the proof of Theorem 8 we use precisely the region $R$ introduced in the proof of Theorem 7, with further restrictions on the numbers $d_k$; the function $F(z)$ is, as before, the mapping function.

It follows from equation (9.3) that the function

$$w = F_k(z) = b_k + \frac{1}{2^k}\left[z + \frac{2}{2}z^2 + \frac{3}{2^2}z^3 + \frac{4}{2^3}z^4 + \cdots\right] \tag{9.8}$$

maps $|z| < 1$ onto the region $J_k$ in such a way that the point $z = 0$ corresponds to the point $B_k$: $w = b_k$ with the axis of reals in one plane corresponding to the axis of reals in the other plane. The function

$$w(\zeta) = F\left(\frac{z_k + \zeta}{1 + z_k\zeta}\right) = b_k + a_1^{(k)}\zeta + a_2^{(k)}\zeta^2 + a_3^{(k)}\zeta^3 + \cdots, \tag{9.9}$$

where $F(z_k) = b_k$, maps $|\zeta| < 1$ onto $R$ so that $\zeta = 0$ corresponds to the point $B_k$ with the axis of reals in the one plane corresponding to the axis of reals in the other. When $d_k$ and $d_{k+1}$ approach zero, the kernel in the sense of Cara-

property (9.2), nor does Sewell, but the latter (these Transactions, vol. 41 (1937), pp. 84–123) mentions for Szegö's region the relation (notation of §1) $\lim_{k \to \infty} D_1(w_k)/Q(|z_k|) = \infty$, $w_k = F(z_k)$, which by virtue of the inequality $|F'(z_n)|(1 - |z_n|^2) \geq D_1(w_n)$ implies (9.2). Szegö's example does not seem to apply at once to higher derivatives.

The method of proof of Theorem 7 has also been employed by Walsh, Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 101–108, for a somewhat different purpose.

théodory[33] of the variable region $R$, considered with $B_k$ as central point (that is, Aufpunkt) is precisely the region $J_k$. It follows from the results of Carathéodory (loc. cit.) that the corresponding mapping function $w(\zeta)$ defined by (9.9) approaches the function $F_k(\zeta)$ defined by (9.8), throughout the interior of $|\zeta|<1$, uniformly on any closed point set interior to $|\zeta|<1$. Indeed, such uniform approach of $w(\zeta)$ defined by (9.9) to $F_k(\zeta)$ is a consequence of the approach to zero of $d_k$ and $d_{k+1}$, *independently of the behavior of* $d_1, d_2, \cdots, d_{k-1}, d_{k+2}, d_{k+3}, \cdots$. Otherwise there would exist a sequence of sequences of numbers $d_1, d_2, \cdots$ with $d_k$ and $d_{k+1}$ approaching zero and the corresponding function $w(\zeta)$ in (9.9) not approaching $F_k(\zeta)$ as defined by (9.8); this is impossible. Thus the coefficient $a_j^{(k)}$, considered as a function of $d_k$ and $d_{k+1}$ alone, approaches the corresponding coefficient $j/2^{j+k-1}$.

The inner radius $\rho(b_k)$ of $R$ with respect to the point $B_k$ is greater than $1/2^k$, so in (9.9) we have

$$a_1^{(k)} > 1/2^k. \tag{9.10}$$

We have already made restrictions on the numbers $d_k$ in connection with Theorem 7. We now impose the further restriction that $d_1, d_2, \cdots$ are to be chosen in pairs $(d_1, d_2), (d_2, d_3), (d_3, d_4), \cdots$ successively so small that we always have the inequalities $(k=1, 2, 3, \cdots)$

$$a_2^{(k)} > 0,\ a_3^{(k)} > 0, \cdots, a_m^{(k)} > 0; \tag{9.11}$$

this choice of the $d_k$ is possible. We have no other restrictions to be placed on the numbers $d_k$.

By Lemma 1 of §2 we now have

$$\frac{(1-|z_k|^2)^m}{m!}F^{(m)}(z_k) = \sum_{\nu=0}^{m-1} C_{m-1,\nu}\bar{z}_k^{\nu}\frac{w^{(m-\nu)}(0)}{(m-\nu)!},$$

where $w(\zeta)$ is defined by (9.9). Inequalities (9.11) and (9.10) now yield $(z_k=\bar{z}_k>0)$

$$\frac{(1-|z_k|^2)^m}{m!}F^{(m)}(z_k) > z_k^{m-1}a_1^{(k)} > \frac{z_k^{m-1}}{2^k}, \qquad k>1,$$

so, as in (9.6), we write from (9.5)

$$\frac{(1-|z_k|^2)^m F^{(m)}(z_k)}{Q(|z_k|)} > m!\cdot z_k^{m-1}\cdot\frac{3^k}{2^k}, \qquad z_k = |z_k|.$$

When $k$ becomes infinite, the point $z_k$ approaches the point $z=1$, so equation (9.7) and Theorem 8 follow.

As will be seen, this function $F(z)$ is significant as a "Gegenbeispiel" also in some of our later theorems.

---

[33] Cf. Footnote 13.

## Chapter II. Bounded functions: configurations $C_p$ and $D_p$

The problem which will occupy us in this chapter and the next is to what extent the results of the first chapter can be extended to the class of bounded functions.

It should be remarked at the start that in §5, Corollary 2, it is not possible to drop the condition of univalence. Indeed we have

**Theorem 1.** *There exists a function $f(z)$ regular and bounded in $|z|<1$ and a sequence of points $z_n$ ($|z_n|<1$), $|z_n|\to 1$, for which*

$$\liminf_{n\to\infty} |f'(z_n)|(1-|z_n|) > 0.$$

That $|f'(z_n)|(1-|z_n|)$ is always bounded when $f(z)$ is regular and bounded in $|z|<1$, follows from an easy application of Schwarz's lemma([34]).

To prove Theorem 1 we consider the function([35])

$$f(z) = \exp\left[\frac{z+1}{z-1}\right].$$

It is clear that since $\Re[(z+1)/(z-1)]<0$ in $|z|<1$, we have $|f(z)|<1$ in $|z|<1$. Now,

$$|f'(re^{i\theta})|(1-r^2) = 2\,\frac{1-r^2}{1-2r\cos\theta+r^2}\cdot\exp\left[\frac{-1+r^2}{1-2r\cos\theta+r^2}\right].$$

Along the curve $r=\cos\theta$ which passes through the point $z=1$ and is tangent there to the unit circle

$$|f'(re^{i\theta})|(1-r^2) = 2/e$$

so that as $\theta\to 0$, the corresponding limit is $2/e>0$.

**10. A lower bound on $D_1(w)$.** In order to obtain the conclusion $\lim_{n\to\infty}|f'(z_n)|(1-|z_n|)=0$, it is necessary to limit oneself to particular sequences $\{z_n\}$ in the circle $|z|<1$. By Theorem 1 our result is as follows:

**Theorem 2.** *Let $f(z)$ be regular and bounded in $|z|<1$:*

$$|f(z)| \leqq M,$$

*let $\{z_n\}$ be any sequence of points in $|z|<1$, and let $w_n=f(z_n)$. Then, a necessary and sufficient condition for*

$$\lim_{n\to\infty} |f'(z_n)|(1-|z_n|) = 0$$

*is that* $\lim_{n\to\infty} D_1(w_n)=0$.

---

([34]) Cf. L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, 2d edition, 1931, p. 112.

([35]) For this particularly simple example the authors are indebted to Professor G. Szegö.

This condition will follow directly from the more precise

THEOREM 3. *Let $f(z)$ be regular and bounded in $|z|<1$:*

$$|f(z)| \leqq M,$$

*let $z_0$ be any point in $|z|<1$, and let $w_0=f(z_0)$. Then, the following inequality*

$$D_1(w_0) \leqq |f'(z_0)|(1-|z_0|^2) \leqq [8MD_1(w_0)]^{1/2} \tag{10.1}$$

*is always satisfied.*

The first inequality in (10.1) is simply a particular case of §4, Theorem 2. It, therefore, remains to prove the second inequality alone.

It was proved by Landau and Dieudonné[38] that if

$$w = g(z) = z + \cdots$$

is a regular function in $|z|<1$ satisfying the inequality

$$|g(z)| \leqq M \quad \text{for} \quad |z| < 1,$$

then $g(z)$ is univalent in the circle $|z|<1/2M$ and covers simply the circle $|w|\leqq 1/4M$.

Consider now the function

$$\phi(z) = \frac{f((z+z_0)/(1+\bar{z}_0 z)) - f(z_0)}{f'(z_0)(1-|z_0|^2)} = z + \cdots .$$

In $|z|<1$ the function $\phi(z)$ is regular and satisfies the inequality

$$|\phi(z)| \leqq \frac{2M}{|f'(z_0)|(1-|z_0|^2)}.$$

Hence, in accordance with the theorem of Landau and Dieudonné $w=\phi(z)$ covers simply the circle

$$|w| \leqq \frac{|f'(z_0)|(1-|z_0|^2)}{8M}.$$

The function $w=f(z)$, therefore, covers simply the circle

$$|w-w_0| \leqq \frac{|f'(z_0)|^2(1-|z_0|^2)^2}{8M}, \qquad w_0 = f(z_0).$$

From this it follows that

---

(38) E. Landau, Sitzungsberichte der Preussischen Akademie der Wissenschaften, Berlin, Physikalisch-Mathematische Klasse, (1926), pp. 467–474; J. Dieudonné, Annales de l'École Normale Supérieure, (3), vol. 48 (1931), pp. 247–358.

$$D_1(w_0) \geqq \frac{|f'(z_0)|^2(1-|z_0|^2)^2}{8M},$$

which is merely another form of the second inequality (10.1)([37]).

While the constant 8 in (10.1) is not the best possible, the order $[D_1(w_0)]^{1/2}$ as $D_1(w)\to 0$ cannot be improved, as may be seen from a study in $|z|<1$ of the function

$$f(z) = \frac{Mz(1-Mz)}{M-z}, \qquad M>1,$$

previously considered by J. Dieudonné([38]) in the neighborhood of the point $z=M-[M^2-1]^{1/2}$. Indeed, let $z$ be any point of the unit circle, lying on the real axis, such that $0<z<M-[M^2-1]^{1/2}$. It is seen by direct computation that

$$|f'(z)|(1-z^2) = M^2\frac{(1-2Mz+z^2)(1-z^2)}{(M-z)^2}$$

$$(10.2) \qquad = \frac{M^2(1-z^2)}{(M-z)^2}[z-(M-(M^2-1)^{1/2})][z-(M+(M^2-1)^{1/2})].$$

We set $w_0=f(M-(M^2-1)^{1/2})=M(M-(M^2-1)^{1/2})^2$. Hence, since $D_1(w)=w_0-w$,

$$(10.3) \qquad D_1(w) = \frac{M^2}{M-z}[z-(M-(M^2-1)^{1/2})]^2.$$

Comparison of the equations (10.2) and (10.3) shows that as $w\to w_0$ $|f'(z)|(1-z^2)=O((D_1(w))^{1/2})$, but $|f'(z)|(1-z^2)\neq o((D_1(w))^{1/2})$.

11. **Irregular sequences.** The question now arises whether one may generalize Theorem 3 to higher derivatives in the same manner as Theorems 4 and 5 generalize Theorem 3 in Chapter I. In the present case, however, the situation is more complicated than in the case of univalent functions, as examples (§12) will show. Before giving the examples it will be desirable to give some definitions and prove two theorems. Being given two points $z_1$ and $z_2$ of the unit circle $|z|<1$, we define as in §7 the non-euclidean distance $\rho(z_1, z_2)$ between them([39]).

DEFINITION 1. *A sequence of points* $\{z_n\}$, $(|z_n|<1)$, $z_n\to 1$, *will be called a regular sequence for a function* $f(z)$ *analytic in* $|z|<1$ *if there exists a number*

([37]) It will be observed from the above that it might be of advantage sometimes to replace the right-hand side of (10.1) by $[4M'D_1(w_0)]^{1/2}$ where $M'$ is the least upper bound of $|f((z+z_0)/(1+\bar{z}_0z))-f(z_0)|$ for $|z|<1$ and $z_0$ fixed.

([38]) J. Dieudonné, ibid.

([39]) For the notions of non-euclidean geometry particularly in their relation to the theory of functions, cf. G. Julia, ***Principes Géométriques d'Analyse***, Première Partie, 1930, especially Chapters II and IV.

*$\lambda > 0$ such that for any sequence of points $\{z_n'\}$ whose non-euclidean distance $\rho(z_n, z_n')$ is less than $\lambda$ for all $n$ we have*

$$\lim_{n\to\infty} [f(z_n) - f(z_n')] = 0.$$

*A sequence of points $\{z_n\}$ which is not regular will be called irregular.*

DEFINITION 2. *A sequence of points $\{z_n\}$, $(|z_n| < 1)$, $z_n \to 1$, will be called a quasi-regular sequence of order $m$ for a function $f(z)$ analytic in $|z| < 1$ if*

$$\lim_{n\to\infty} |f^{(k)}(z_n)| (1 - |z_n|)^k = 0, \quad \textit{for } k = 1, 2, \cdots, m,$$

*while*

$$\limsup_{n\to\infty} |f^{(m+1)}(z_n)| (1 - |z_n|)^{m+1} > 0.$$

*The case $m = \infty$ is allowed and means that* $\lim_{n\to\infty} |f^{(k)}(z_n)| (1-|z_n|)^k = 0$ for $k = 1, 2, \cdots$.

Denote by $\Gamma_n^\lambda$ the non-euclidean circle of non-euclidean radius $\lambda$ and non-euclidean center $z_n$. We prove now the following

THEOREM 4. *An irregular sequence $\{z_n\}$ for a function $f(z)$ regular and bounded in $|z| < 1$ is quasi-regular of order $m$ if to every sufficiently small positive $\lambda$ there corresponds an integer $N(\lambda) > 0$ such that for all $n > N(\lambda)$ the function $f(z)$ assumes the value $f(z_n)$ exactly $m+1$ times in the circle $\Gamma_n^\lambda$ (counting multiplicities).*

Consider the function

$$g_n(\zeta) = f\left(\frac{\zeta + z_n}{1 + \bar{z}_n\zeta}\right) - f(z_n). \tag{11.1}$$

By hypothesis, for $n > N(\lambda)$ the function $g_n(\zeta)$, which is regular and bounded in $|\zeta| < 1$, assumes the value 0 exactly $m+1$ times in the circle $|\zeta| < (e^\lambda - 1)/(e^\lambda + 1)$. Now, the sequence $\{z_n\}$ is assumed to be irregular. In accordance with Definition 1 this means that for any $\lambda > 0$ we can find a subsequence of the $\{z_n\}$, which we shall denote by $\{z_{n_k}\}$, and a sequence $\{z'_{n_k}\}$ such that $\rho(z_{n_k}, z'_{n_k}) < \lambda$ and for some $\delta > 0$ we have $|f(z_{n_k}) - f(z'_{n_k})| \geqq \delta$. This implies, however, that the sequence (11.1) cannot tend uniformly to zero in every closed subregion of $|\zeta| < 1$. Indeed, suppose that $\lim_{n\to\infty} g_n(\zeta) = 0$ uniformly in every closed subregion of $|\zeta| < 1$. To any preassigned $\epsilon > 0$ there would correspond a positive integer $n(\epsilon)$ so that for $n > n(\epsilon)$ we would have $|g_n(\zeta)| < \epsilon$ in $|\zeta| < (e^\lambda - 1)/(e^\lambda + 1)$. Setting $\zeta_{n_k} = (z'_{n_k} - z_{n_k})/(1 - \bar{z}_{n_k} z'_{n_k})$ we would infer that $|g_{n_k}(\zeta_{n_k})| < \epsilon$ for $n > n(\epsilon)$. Replacing this inequality in (11.1), we find $|f(z'_{n_k}) - f(z_{n_k})| < \epsilon$ for $n > n(\epsilon)$. If we choose $\epsilon < \delta$, we arrive at a contradiction.

Hence, there exists[40] a subsequence of the sequence $\{g_n(\zeta)\}$, which we shall denote by $\{g_{n_k}(\zeta)\}$, which converges uniformly in every closed subregion of $|\zeta|<1$ to a function $G(\zeta)$ which is not identically zero, and (since $G(0)=0$) is not identically a constant. The function $G(\zeta)$ is regular in $|\zeta|<1$.

Since $G(\zeta)$ is not identically zero, there must exist a $0<\lambda_1<\lambda$ so that $G(\zeta)\neq 0$ on the circle $|\zeta|=(e^{\lambda_1}-1)/(e^{\lambda_1}+1)$. Since, furthermore, the sequence $g_{n_k}(\zeta)$ converges uniformly to $G(\zeta)$ on that circle, for sufficiently large values of $n_k$ we have $g_{n_k}(\zeta)\neq 0$ on $|\zeta|=(e^{\lambda_1}-1)/(e^{\lambda_1}+1)$. Now, by hypothesis $g_{n_k}(\zeta)$ vanishes precisely $m+1$ times in the circle $|\zeta|<(e^{\lambda_1}-1)/(e^{\lambda_1}+1)$ provided $n_k>N(\lambda_1)$. Hence, by Hurwitz's theorem $G(\zeta)$ vanishes precisely $m+1$ times in the circle $|\zeta|<(e^{\lambda_1}-1)/(e^{\lambda_1}+1)$. But since $\lambda$ may be taken arbitrarily small, $G(\zeta)$ must have a zero of order $m+1$ at the origin. Hence, $G'(0)=0$, $G''(0)=0,\cdots, G^{(m)}(0)=0$, $G^{(m+1)}(0)\neq 0$. In view of (11.1) and (2.2), we see from the relations $g'_{n_k}(0)\to 0$, $g''_{n_k}(0)\to 0,\cdots, g^{(m)}_{n_k}(0)\to 0$, $g^{(m+1)}_{n_k}(0)\to G^{(m+1)}(0)$ that

$$\limsup_{n\to\infty} |f^{(m+1)}(z_n)|(1-|z_n|)^{m+1}>0.$$

On the other hand, suppose that for some integer $0<k<m+1$ and for some subsequence $\{z_{n'}\}$ of the $\{z_n\}$

$$|f^{(k)}(z_{n'})|(1-|z_{n'}|)^k>\eta>0 \tag{11.2}$$

for a suitable positive $\eta$, independent of $n'$. Consider the corresponding subsequence $\{g_{n'}(\zeta)\}$ of the sequence (11.1). By selecting a further subsequence, if necessary, we may assume that the sequence $\{g_{n'}(\zeta)\}$ is a uniformly convergent one in every closed subregion of $|\zeta|<1$. Two cases are possible according as $\{g_{n'}(\zeta)\}$ converges to zero or to some function not identically a constant. In the first case, the derivatives of all orders of $g_{n'}(\zeta)$ also converge to zero and application of formula (2.2) for the case $n=k$ shows that $|f^{(k)}(z_{n'})|\ (1-|z_{n'}|)^k\to 0$, which is a contradiction of (11.2). In the second case, the nonconstant limit function $G(\zeta)$ of the sequence $g_{n'}(\zeta)$ by the argument already given must have a zero of order $m+1$ at the origin, so that all its derivatives up to the $(m+1)$st must vanish at the origin. Application of formula (2.2) again contradicts (11.2). Thus, in both cases (11.2) yields a contradiction. Hence,

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k=0, \qquad k=1,2,\cdots,m,$$

and the sequence $\{z_n\}$ is quasi-regular of order $m$.

We proceed to prove some related results.

---

(40) This follows from the fact that the functions $g_n(\zeta)$, being uniformly bounded in their totality in $|\zeta|<1$, form a normal family, cf. P. Montel, *Leçons Sur les Familles Normales de Fonctions Analytiques*, 1927, p. 21.

THEOREM 5. *A necessary and sufficient condition that $\{z_n\}$ be a regular sequence for a function $f(z)$ regular in $|z|<1$ and bounded there: $|f(z)|\leqq M$, is that it be quasi-regular of infinite order.*

The condition is necessary. Indeed, form the functions

$$g_n(\zeta) = f\left(\frac{\zeta + z_n}{1 + \bar{z}_n\zeta}\right) - f(z_n). \tag{11.3}$$

This sequence of functions is uniformly bounded in $|\zeta|<1$. From every subsequence can be extracted a new subsequence whose uniform limit is zero in every circle $\rho(\zeta,0)<\lambda$, where $\lambda$ is the number of Definition 1. It follows that the sequence $g_n(\zeta)$ converges uniformly to zero in the circle $|\zeta|\leqq(e^\rho-1)/(e^\rho+1)$, $\rho<\lambda$. Lemma 1 of §2 shows that $|f^{(k)}(z_n)|\,(1-|z_n|)^k\to 0$ for $k=1, 2, \cdots$.

The condition is sufficient. Again form the functions (11.3). Since $|g_n(\zeta)|\leqq 2M$ in $|\zeta|<1$, the functions form a normal family. A suitable subsequence converges uniformly in every circle $|\zeta|\leqq d<1$ to some function $G(\zeta)$ which is regular and bounded in $|\zeta|<1$ and $G(0)=0$. Expanding $G(\zeta)$ in a Taylor series about $\zeta=0$:

$$G(\zeta) = c_1\zeta + c_2\zeta^2 + \cdots, \tag{11.4}$$

applying Lemma 2 of §2 and the hypothesis that $\{z_n\}$ is quasi-regular of infinite order, we see that all the coefficients in the expansion (11.4) are zero and that therefore $G(\zeta)\equiv 0$.

Since we may repeat this argument starting with any subsequence of the family $\{g_n(\zeta)\}$, it follows that $g_n(\zeta)\to 0$ uniformly in any circle $|\zeta|\leqq d<1$. From this follows at once the fact that $\{z_n\}$ is a regular sequence for $f(z)$.

A type of converse of Theorem 4 may be stated in the following form:

THEOREM 6. *Let $f(z)$ be regular and bounded in the unit circle $|z|<1$: $|f(z)|\leqq M$. Let the sequence $\{z_n\}$ be quasi-regular of order $m$. Then for every subsequence of the $\{z_n\}$ there exists a new subsequence $\{z_{n_k}\}$ with the property that to every $\rho>0$ which is sufficiently small there corresponds an integer $N(\rho)>0$ such that for all $n_k>N(\rho)$ the function $f(z)$ assumes the value $f(z_{n_k})$ precisely $m+1$ times in the circle $\Gamma_n^\rho$ (counting multiplicities).*

Again we form the functions (11.3). In view of Lemma 2 of §2 we have

$$\frac{g_n^{(p)}(0)}{p!} = \sum_{\nu=0}^{p-1}(-1)^\nu C_{p-1,\nu}\bar{z}_n^\nu\frac{(1-|z_n|^2)^{p-\nu}f^{(p-\nu)}(z_n)}{(p-\nu)!}.$$

The hypothesis that $\{z_n\}$ is a quasi-regular sequence of order $m$ for $f(z)$ implies that

$$\lim_{n\to\infty} g_n^{(p)}(0) = 0, \qquad \text{for } p = 1, 2, \cdots, m, \tag{11.5}$$

while

$$\limsup_{n\to\infty} |g_n^{(m+1)}(0)| > 0. \tag{11.6}$$

Let us select a subsequence of the family $\{g_n(\zeta)\}$ for which the actual limit in (11.6) exists and is positive and denote this subsequence for simplicity by $\{g_n(\zeta)\}$ again. Since for all $n$ we have $|g_n(\zeta)| \leqq 2M$ in $|\zeta| < 1$, the sequence $\{g_n(\zeta)\}$ is a normal family. We may therefore extract a further subsequence $\{g_{n_k}(\zeta)\}$ which in every closed subregion of the circle $|\zeta| < 1$ converges uniformly to a function $G(\zeta)$. According to (11.5) and (11.6) we obtain $G^{(p)}(0) = 0$ for $p = 1, 2, \cdots, m$ and $G^{(m+1)}(0) \neq 0$. Since $G(0) = 0$, it follows that for every $\rho > 0$ which is sufficiently small the function $G(\zeta)$ has precisely $m+1$ zeros in the circle $|\zeta| < \rho$ and is different from zero on the circumference $|\zeta| = \rho$. Let us fix a definite value of $\rho$. By Hurwitz's theorem it follows that there exists an integer $N(\rho) > 0$ so that each function $g_{n_k}(\zeta)$ for which $n_k > N(\rho)$ has precisely $m+1$ zeros in the circle $|\zeta| < \rho$. The theorem then follows immediately from the definition (11.3) of $g_n(\zeta)$.

12. **Counterexamples (Gegenbeispiele).** Theorem 4 may be used to obtain an example in which $D_1(w_n) = 0$, while $|f''(z_n)|\ (1 - |z_n|)^2$ does not tend to zero. Indeed, consider the Blaschke product[41]

$$\phi(z) = \prod_{n=1}^{\infty} z_n \frac{1 - z/z_n}{1 - z_n z}, \qquad z_n = \frac{n! - 1}{n! + 1}. \tag{12.1}$$

As is well known[42], since $\prod_{n=1}^{\infty}(n!-1)/(n!+1)$ converges, the product (12.1) represents in the circle $|z| < 1$ an analytic function whose absolute value is less than unity. As was shown by one of the authors[43], the sequence $\{z_n\}$ is an irregular sequence for $\phi(z)$. Now, form

$$f(z) = [\phi(z)]^2.$$

Again, the sequence $\{z_n\}$ is an irregular sequence for $f(z)$. Furthermore, since the $z_n$ are zeros of order 2 and the only zeros of $f(z)$, we have $D_1(0) = 0$ when the point $w = 0$ is considered in any sheet of the Riemann configuration for $w = f(z)$. On the other hand, the non-euclidean distance $\rho(z_n, z_{n+1}) = \log(n+1) \to \infty$. Hence, for any $\lambda > 0$ and for sufficiently large values of $n$ the function $f(z)$ vanishes precisely twice in $\Gamma_n^\lambda$. Applying Theorem 4, therefore, we find

$$\limsup_{n\to\infty} |f''(z_n)|\ (1 - |z_n|)^2 > 0.$$

---

(41) Such products were first introduced by W. Blaschke, Berichte über die Verhandlungen der Sächsischen Akademie der Wissenschaften, Mathematisch-Physische Klasse, Leipzig, vol. 67 (1915), pp. 194–200.

(42) Cf. G. Julia, ibid., pp. 65–66.

(43) W. Seidel, these Transactions, vol. 34 (1932), pp. 14–15. Equation (7.2) there should read

$$\phi(z) = \prod_{n=1}^{\infty} t_n \frac{1 - z/t_n}{1 - t_n z}.$$

Let us state this example as a theorem:

THEOREM 7. *There exists a bounded regular function* $f(z)$ *in* $|z|<1$ *and a sequence of points* $\{z_n\}$ $(|z_n|<1, |z_n|\to 1)$ *in* $|z|<1$ *such that, setting* $w_n=f(z_n)$, $\lim_{n\to\infty} D_1(w_n)=0$, *while* $\limsup_{n\to\infty} |f''(z_n)|(1-|z_n|)^2>0$.

Indeed, for the specific example already given we may assert $f'(z_n)=0$, $D_1(w_n)=0$.

The converse situation may also arise:

THEOREM 8. *There exists a bounded regular function* $f(z)$ *in* $|z|<1$ *and a sequence of points* $\{z_n\}$ $(|z_n|<1, |z_n|\to 1)$ *in* $|z|<1$ *such that, setting* $w_n=f(z_n)$, *we have* $\liminf_{n\to\infty} D_1(w_n)>0$, *while* $\lim_{n\to\infty} |f''(z_n)|(1-|z_n|)^2=0$.

Let

$$\phi(z) = (1 - e^{-W+1})^2, \qquad W = \frac{1+z}{1-z}\cdot$$

The function $W=(1+z)/(1-z)$ maps the circle $|z|<1$ on the half-plane $\Re W>0$. Now, for $\Re W>0$

$$|e^{-W+1}| = \exp(-\Re W + 1) < e.$$

Hence, for $|z|<1$, $|\phi(z)|<(1+e)^2$.

Direct computation shows that

$$\phi'(z) = -4e^{-W+1}(1 - e^{-W+1})\cdot\frac{1}{(1-z)^2},$$

$$\phi''(z) = \frac{8}{(1-z)^3}e^{-W+1}(e^{-W+1} - 1) + \frac{8}{(1-z)^4}e^{-W+1}(1 - 2e^{-W+1}).$$

We now choose the points $z_n=n\pi i/(n\pi i+1)$. It is clear that $|z_n|<1$ and $\lim_{n\to\infty} z_n=1$. Setting $W_n=(1+z_n)/(1-z_n)$, we find $W_n=1+2n\pi i$. Hence

$$(12.2) \qquad |\phi'(z_n)|(1 - |z_n|) = 0, \quad \phi''(z_n)(1 - |z_n|^2)^2 \to 8 \quad \text{as} \quad n\to\infty.$$

This incidentally gives another example for the proof of Theorem 7, since the relation $D_1(w_n)\to 0$ follows from (12.2) and (10.1).

Next, we introduce the function

$$\psi(z) = e^{-W+1}, \qquad W = \frac{1+z}{1-z}\cdot$$

Again, we observe that in $|z|<1$ the function $\psi(z)$ is bounded:

$$|\psi(z)| < e.$$

Choosing again $z_n=n\pi i/(n\pi i+1)$, we find

$$(12.3) \qquad |\psi'(z_n)|(1 - |z_n|^2) = 2, \qquad \text{for all } n,$$

and

$$\psi''(z_n)\cdot(1-|z_n|^2)^2 \to 4 \quad \text{as} \quad n\to\infty. \tag{12.4}$$

Finally, we introduce the function

$$f(z) = \phi(z) + 2\psi(z).$$

It is clear that $f(z)$ is bounded in the circle $|z|<1$, satisfying there the inequality

$$|f(z)| < (1+e)^2 + 2e.$$

For the sequence of points $z_n = n\pi i/(n\pi i+1)$ by virtue of (12.2), (12.3), (12.4) we have the relations

$$|f'(z_n)|\,(1-|z_n|^2) = 4 \neq 0, \qquad \text{for all } n,$$

and

$$f''(z_n)(1-|z_n|^2)^2 \to 0 \quad \text{as} \quad n\to\infty.$$

It follows from Theorem 3 that $D_1(w_n)$ has a positive lower bound. The function $f(z)$ is therefore an example of a function with the properties asserted in Theorem 8, and Theorem 8 is established.

By forming the function $f(z)=a\phi(z)+b\psi(z)$ with arbitrary constants $a$ and $b$, one can now obtain arbitrary limits for $|f'(z_n)|\ (1-|z_n|^2)$ and $|f''(z_n)|\,(1-|z_n|^2)^2$ as $n\to\infty$.

It may be observed that if the condition $|z_n|\to 1$ in Theorems 7 and 8 were dropped, one might take as examples to prove the theorems the simple functions $f(z)=z^2$ and $f(z)=z$, $z_n=0$, respectively.

Finally, it may be noted that in the Theorems 4–6, the boundedness of $f(z)$ was assumed merely in order to ensure the normality of the family $g_n(\zeta)$. Thus, it would have sufficed to assume that $f(z)$ has 2 exceptional values and $f(z_n)$ is bounded.

## 13. Definition and some properties of $C_p$.

DEFINITION 3. *Let $C_p$ be a simply connected Riemann configuration containing the point $w_0$, lying over the circle $|w-w_0|<\rho$ and covering it precisely $p$ times. Such a region $C_p$ will be called a $p$-sheeted circle of center $w_0$ and radius $\rho$.*

We shall exclude the case $\rho=\infty$ (called an improper $p$-sheeted circle) for a reason that will be given a little later. It should be observed that the center of a $p$-sheeted circle is not uniquely defined.

The necessity of assuming explicitly (rather than proving) in Definition 3 that $C_p$ shall be simply connected may be seen from the following example. Consider in the $w$-plane the (simply connected) Riemann surface of the function $((w-\alpha)/(w-\beta))^{1/2}$ where $\alpha$ and $\beta$ are two complex numbers, with the branch line the rectilinear segment $\alpha\beta$. Let us now cut this surface by a circular biscuit-cutter which includes the two points $\alpha$ and $\beta$. The resulting circular

region cut out of the surface satisfies all the requirements in Definition 3 except the condition of simple connectivity. In fact, every region lying over a circle $|w-w_0|<\rho$ and covering it precisely twice ceases to be simply connected as soon as it has two branch points or more. Indeed, in such a case it is clearly possible to find a cut joining two boundary points and crossing a branch line which will not sever the surface. In general, by applying the theorem of Bôcher and Walsh (as in the proof of Theorem 13 below) one may easily show that every region lying over a circle $|w-w_0|<\rho$ and covering it precisely $p$ times ceases to be simply connected as soon as the sum of the multiplicities of its branch points exceeds $p-1$. The multiplicity of a branch point is to be understood as one less than the number of sheets which come together at that point. An algebraic branch point (but not a transcendental one) is to be considered as belonging to the Riemann configuration.

One may prove some immediate consequences of Definition 3.

THEOREM 9. *Any $p$-sheeted circle over the $w$-plane can be mapped in a one-to-one conformal manner on the unit circle* $|z|<1$.

According to the fundamental theorem of uniformization the $p$-sheeted circle $C_p$, being simply connected, may be mapped in a one-to-one conformal manner either on a circle, or on the full plane, or on the full plane from which the point at infinity is excluded. Denote the mapping function by $w=f(z)$. Since $C_p$ is a bounded region, the function $f(z)$ must be bounded. This is certainly not possible in the two latter cases. Thus, $C_p$ can be mapped only on a circle.

THEOREM 10. *A $p$-sheeted circle $C_p$ with center $w_0$ and radius $\rho$ can be mapped in a one-to-one conformal manner on the unit circle $|z|<1$ by means of a function of the form*

$$f(z) = w_0 + \rho e^{i\theta} z^k \prod_{j=1}^{p-k} \frac{z - z_j}{1 - \bar{z}_j z}, \tag{13.1}$$

*where $\theta$ is an arbitrary real number, $k$ an integer satisfying the inequality $0<k\leqq p$, and where $z_1, z_2, \cdots, z_{p-k}$ are points of the unit circle $|z|<1$. Conversely, every function of the form* (13.1) *realizes a one-to-one conformal map of the unit circle $|z|<1$ on some $p$-sheeted circle with center at $w_0$ and radius $\rho$.*

In speaking of conformality, it must be remembered that it will break down at a branch point. To prove the first part of the theorem introduce a similarity transformation in the $w$-plane with center in $w_0$ which transforms the circle $C_p$ into a $p$-sheeted circle $C_p'$ of radius 1. By means of a translation we can always bring the point $w_0$ into the origin. The resulting one-to-one map of $|z|<1$ on $C_p'$ can be interpreted as a $(1, p)$ conformal corre-

(44) T. Radó, Acta Litterarum ac Scientiarum Regiae Universitatis Hungaricae Francisco-Josephinae, Szeged, vol. 1 (1922), p. 55. Preliminary related work is due also to Fatou and Julia.

spondence of a unit circle on itself. By applying Radó's theorem([44]) on the representation of such correspondences we obtain the expression (13.1). The converse may also be derived from Radó's theorem together with a translation and similarity transformation in the $w$-plane.

A remark will now be made to justify the exclusion of the case $\rho=\infty$ in the definition of $C_p$. An improper $p$-sheeted circle could be interpreted as the $w$-plane covered precisely $p$ times. If such a circle belonged to a simply connected Riemann surface, the surface could not be of hyperbolic type and consequently Theorems 9 and 10 would no longer apply. Suppose first that a simply connected Riemann surface which contains an improper $p$-sheeted circle could be mapped conformally on the unit circle. Thereby the $p$-sheeted circle would be transformed into a simply connected subregion of the unit circle. Now if the improper $p$-sheeted circle has no boundary points such a transformation is clearly impossible. Suppose then that the $p$-sheeted circle has the point $w=\infty$ as a boundary point. Then, the mapping function in the unit circle approaches infinity whenever the point $z$ approaches the boundary of the subregion. This is again impossible.

THEOREM 11. *Let $C_p$ be a $p$-sheeted circle with center at $w_0$ and radius $R$. Let $c_p$ be a subregion of $C_p$ which lies over a circle $|w-w_0|<r$, where $r<R$, and covers it precisely $p$ times. Then, $c_p$ is also simply connected.*

We can map $C_p$ on the unit circle $|z|<1$ in accordance with Theorem 9. The mapping function $w=f(z)$ is regular in $|z|<1$ and maps $c_p$ on a certain subregion $B$ of $|z|<1$. On the boundary $\Gamma$ of $B$ we have $|f(z)-w_0|=r$, while in the interior of $B$ we have $|f(z)-w_0|<r$. From the maximum modulus principle it follows that $B$ is simply connected. Since the map defined by $w=f(z)$ is topological, the image of $c_p$ of $B$ must likewise be simply connected.

In order to establish the uniqueness in Definition 3, we shall prove

THEOREM 12. *Let $R$ be a simply connected Riemann surface of hyperbolic type. Let $w_0$ be a point of $R$. Let $C_p$ and $C_p'$ be two $p$-sheeted circles with center at $w_0$ and radius $\rho$. Then, $C_p$ and $C_p'$ are identical.*

If we map $R$ on the unit circle $|z|<1$ by means of the function $w=f(z)$ so that $f(0)=w_0$, the two circles $C_p$ and $C_p'$ will be mapped on two regions $B$ and $B'$ belonging to the circle $|z|<1$. In the interiors of $B$ and $B'$ we have $|f(z)-w_0|<\rho$ and on the boundaries $|f(z)-w_0|=\rho$. Furthermore, both regions $B$ and $B'$ contain the origin. Thus, unless $B$ and $B'$ are identical at least one boundary point of one region, say $B$, will be interior to the other region $B'$. This, however, constitutes a contradiction.

14. **Definition of $D_p$.**

DEFINITION. *Let $w=f(z)$ regular in the unit circle $|z|<1$ map the circle on a Riemann configuration $R$. That is to say, $R$ is an arbitrary simply connected Riemann configuration of hyperbolic type over the finite $w$-plane. Let $w_0$ be an*

*arbitrary point belonging to R. A non-negative number $D_p(w_0)$, called the radius of p-valence of R at the point $w_0$, shall be associated with the point $w_0$ in the following manner:*

(a) *For $p=1$, we define $D_p(w_0)=D_1(w_0)$ (see §1).*

(b) *If there exists a p-sheeted circle with center $w_0$ contained in R, there exists a largest such circle, and the radius of this largest circle is defined as $D_p(w_0)$.*

(c) *If $p>1$, and if $w_0$ is a branch point of order greater than $p-1$, then $D_p(w_0)=0$.*

(d) *If there exists no p-sheeted circle ($p>1$) with center $w_0$ contained in R, and if $w_0$ is not a branch point of order greater than $p-1$, then we define $D_p(w_0)$ as $D_{p-1}(w_0)$.*

It should be observed that in the definition in part (b) the existence of a largest $p$-sheeted circle with center in $w_0$ contained in $R$ is asserted and still requires some justification. From Theorem 12 it follows that if such a circle exists, it must be unique. Furthermore, as one starts with a $p$-sheeted circle with center in $w_0$ contained in $R$ and proceeds to enlarge its radius, it can never happen that it becomes multiply connected and on enlarging the radius still more, finally again becomes simply connected. This possibility is ruled out by Theorem 11. Finally, the existence of a $p$-sheeted circle with center in $w_0$ and contained in $R$ whose radius is the least upper bound of the radii of all $p$-sheeted circles with center in $w_0$ and contained in $R$ can be established by simple considerations of continuity, which are left to the reader.

The number $D_p(w_0)$ is not, as the notation would seem to indicate, a function merely of $w_0$, a value of $w$, but is rather a function of a specific point of $R$ whose affix is $w_0$; thus $D_p(w_0)$ is precisely a function of $z_0$, where $R$ is determined by the transformation $w=f(z)$. However, no confusion is likely to result from the slight lack of definiteness in the notation $D_p(w_0)$. We denote by $R_p(w_0)$ the unique region of $R$ which is a $q$-sheeted circle $C_q$ ($q\leqq p$) whose center is $w_0$ and radius $D_p(w_0)$.

For the sake of clearness, we present now a numerical illustration of the definition of $D_p(w_0)$. Let $R$ consist of the doubly-carpeted unit circle $|w|<1$ with branch point of the first order at the origin $w=0$, except that in the second sheet there is deleted the subregion of $|w|<1$ contained in the region $|w+1|<1/3$; for definiteness choose the branch line as the segment $0\leqq w<1$; of course this configuration $R$ can be mapped in a one-to-one manner on $|z|<1$ by a single-valued function $w=f(z)$, as can be seen at once by use of the auxiliary transformation $w=z_1^2$, which maps $R$ onto a smooth Jordan region of the $z_1$-plane. We obviously have $D_2(0)=2/3$, for the doubly-carpeted (that is, two-sheeted) circle $|w|<2/3$ is contained in $R$, and that is true of no larger concentric doubly-carpeted circle. When $w_0$ is positive, and in either sheet of $R$, we have

$$D_2(w_0) = w_0 + 2/3, \qquad 0 \leqq w_0 \leqq 1/6,$$

$$D_2(w_0) = 1 - w_0, \qquad 1/6 \leqq w_0 < 1;$$

for positive $w_0$, the size of the region $R_2(w_0)$ is limited by the nearer of the two points $-2/3$, $+1$. When $w_0$ moves from the origin to the left in the first sheet, the size of $R_2(w_0)$ continues to be limited by the point $w=-2/3$:

$$D_2(w_0) = w_0 + 2/3, \qquad -1/3 \leqq w_0 < 0.$$

But when the point $w_0$ continues to the left from the point $w_0=-1/3$, the size of $R_2(w_0)$ is now no longer limited by the point $-2/3$, but is conditioned by the necessity of including no point of $|w+1|<1/3$, hence is limited by the origin; the corresponding region cut out of $R$ is smooth, merely the region $|w-w_0|<|w_0|$:

$$D_2(w_0) = -w_0, \qquad -1/2 \leqq w_0 \leqq -1/3.$$

As $w_0$ moves further to the left from $w=-1/2$, still in the first sheet of $R$, the region $R_2(w_0)$ is now limited only by the point $w=-1$:

$$D_2(w_0) = 1 + w_0, \qquad -1 < w_0 \leqq -1/2.$$

When $w_0$ moves from the origin to the left in the second sheet of $R$, the size of $R_2(w_0)$ is also limited by the point $w=-2/3$:

$$D_2(w_0) = w_0 + 2/3, \qquad -2/3 < w < 0;$$

this situation continues as $w_0$ moves from the value zero to the value $-2/3$, but the region $R_2(w_0)$ is a doubly-carpeted circle for $-1/3<w<0$, and is singly-carpeted (smooth) for $-2/3<w\leqq -1/3$. This completes the study of our numerical case.

Let us now discuss the manner in which $D_p(w_0)$ and $R_p(w_0)$ vary on the general Riemann configuration $R$, the image of $|z|<1$ under the arbitrary map $w=f(z)$, where $f(z)$ is analytic in $|z|<1$. The various possibilities that arise are illustrated by the example just given. We cut all the sheets of $R$ through with a circular biscuit-cutter whose center is $w_0$ and whose radius is the variable $r$. One of the connected sets thus cut out of $R$ contains $w_0$ and is denoted by $R_1$. When $r$ is small it follows from the usual implicit function theorem that if $w_0$ is not a branch point of $R$ the region $R_1$ is smooth, and if $w_0$ is a $q$-fold point of $R$, then $R_1$ consists of a $q$-sheeted circle whose only branch point is $w_0$. As $r$ is gradually increased, this situation continues until the boundary of $R_1$ reaches either a boundary point of $R$ or a branch point of $R$. In the former case we have $D_p(w_0)$ equal to this particular value $r_1$ of $r$, and $R_1$ is $R_p(w_0)$. In the latter case if $r$ is further increased, it may be that $R_1$ becomes a $q'$-sheeted circle with $q<q'\leqq p$, in which case we have $D_p(w_0)\geqq r>r_1$. But it may occur that whenever $r$ is near to but greater than $r_1$ the region $R_1$ is a $q''$-sheeted circle, $q''>p$, in which case we have $D_p(w_0)=r_1$; it may also occur that whenever $r$ is near to but greater than $r_1$ the region $R_1$ has boundary points in common with $R$, in which case we have also $D_p(w_0)=r_1$. If we

have $D_p(w_0) > r_1$, the radius $r$ can be perhaps increased until still further branch points of $R$ lie interior to $R_1$, while $R_1$ remains a $q_1$-sheeted circle whose center is $w_0$, with $q_1 \leq p$. In any case the radius $r$ can be increased from zero to such a value $r_2$ that: (i) either a boundary point of $R$ lies on the boundary of $R_1$, (ii) or there lie on the boundary of $R_1$ branch points of $R$ of such multiplicities that for all values of $r$ slightly greater than $r_2$ the region $R_2$ containing $w_0$ and cut out of $R$ by the biscuit-cutter with center $w_0$ and radius $r$ is a $q''$-sheeted circle with $q'' > p$, (iii) or there lie on the boundary of $R_1$ branch points of $R$ of such nature that for all values of $r$ slightly greater than $r_2$ this region $R_2$ has boundary points which satisfy the relation $|w - w_0| < r_2$. It is to be noted that if the biscuit-cutter of radius $r$ cuts from $R$ the region $R_1$ containing $w_0$, and if $R_1$ has a boundary point $w_1$ (necessarily a boundary point of $R$) for which $|w_1 - w_0| < r$, then we must have $D_p(w_0) < r$. For under these conditions $R_1$ cannot be a $q$-sheeted circle; the point $w_1$ of the $w$-plane may be covered by $R_1$ precisely $q$ times (not necessarily by $q$ sheets meeting at $w_1$), but then (by the implicit function theorem) a suitably chosen neighborhood of $w_1$ is also covered precisely $q$ times by the sheets of $R_1$ that cover $w_1$, and suitable points $w$ in this neighborhood are covered more than $q$ times in all, for they are covered also by $R_1$ in the neighborhood of the boundary point $w_1$.

It is of interest to trace also the situation in the $z$-plane corresponding to the preceding discussion. When $r$ is sufficiently small, $r > 0$, the locus $|f(z) - w_0| = r$ consists (in addition to possible other arcs or curves) of a Jordan curve $J(r)$ in the neighborhood of the point $z_0$, where $w_0 = f(z_0)$; for $r$ sufficiently small, interior to $J(r)$ the function $f(z)$ takes on every value that it assumes (by Theorem 13 below) precisely a number of times equal to the multiplicity $q$ of $z_0$ as a zero of the function $f(z) - w_0$; the image of the interior of $J(r)$ over the $w$-plane is a $q$-sheeted circle of radius $r$ whose only branch point is $w = w_0$. As $r$ now increases, this situation continues until $J(r)$ reaches $|z| = 1$ or until at least one multiple point of $J(r)$ appears (at a multiple point the tangents to $J(r)$ are equally spaced); in the former case we simply have $D_p(w_0)$ equal to the corresponding value $r_1$ of $r$; in the latter case for values of $r$ near to but slightly greater than $r_1$, the locus $|f(z) - w_0| = r$ consists of a Jordan arc $J_1$ near but exterior to $J(r_1)$ plus other Jordan arcs forming with $J_1$ a maximal connected set which we denote by $J(r)$; still other Jordan arcs may belong to the locus and not be connected with $J_1$, but such arcs do not concern us at present. If for every $r$ near to but slightly greater than $r_1$ the set $J(r)$ has a boundary point on $|z| = 1$, then we have $D_p(w_0) = r_1$; in the contrary case $J(r)$ consists of a Jordan curve in $|z| < 1$ containing $J(r_1)$ in its interior; the function $f(z)$ takes on interior to $J(r)$ all the values that it takes on there the same number of times, say $q'$. If $q'$ is greater than $p$ we have $D_p(w_0) = r_1$, but if $q'$ is not greater than $p$ we have $D_p(w_0) > r_1$, and the process of enlarging $J(r)$ can continue beyond $r = r_1$. The process continues as $r$ increases, and $J(r)$ may pass through multiple points, thereby increasing

not merely $r$ but also the number of times (the same for all values) that $f(z)$ takes on interior to $J(r)$ values that it takes on there. The process eventually comes to an end at some value $r=r_2=D_p(w_0)$, either because $J(r_2)$ reaches the boundary $|z|=1$ and hence is no longer a Jordan curve in $|z|<1$, or because the locus $|f(z)-w_0|=r_2$ has a multiple point, and for every $r>r_2$ but near to $r_2$ the locus $|f(z)-w_0|=r$ either fails now to separate $z_0$ from $|z|=1$ or divides $|z|<1$ into regions of which the one containing $z_0$ is a Jordan region in which each value assumed is assumed more than $p$ times.

15. **Some properties of $D_p$.** We return now to the general theory of $D_p(w_0)$; an important tool is([45])

THEOREM 13. *Let $f(z)$ not identically constant be analytic in the simply connected region $B$, let $|f(z)|$ be continuous in the corresponding closed region and have the constant value $b$ on the boundary $C$ of $B$. Then all values $w$ taken on by $f(z)$ in $B$ are taken on there the same number of times $q$, and $f'(z)$ has precisely $q-1$ zeros interior to $B$.*

The region $B$ cannot be the entire plane or the entire plane with the omission of a single point, so $B$ can be mapped conformally onto the interior of the unit circle $\gamma$. It is sufficient to establish the theorem where $B$ is the interior of $\gamma$, which we shall now do. We must have $b>0$, so by the well known properties of the maxima and minima of $|f(z)|$, the zeros of $f(z)$ interior to $\gamma$ are finite in number, $\beta_1, \beta_2, \cdots, \beta_q$ with $q>0$. The function

$$f(z)\cdot\prod_{k=1}^{q}\frac{1-\bar{\beta}_k z}{z-\beta_k},$$

when suitably defined in the points $\beta_k$, is analytic and different from zero at every point interior to $\gamma$; its modulus is continuous in the corresponding closed region and takes the constant value $b$ on $\gamma$. Hence this function itself is a constant of modulus $b$, and we have

$$f(z) = \omega b\prod_{k=1}^{q}\frac{z-\beta_k}{1-\bar{\beta}_k z}, \qquad |\omega| = 1.$$

The first part of Theorem 13 now follows from Rouché's theorem, for if we have $|c|<b$ we have on $\gamma$ the inequality $|c|<|f(z)|$. The latter part of Theorem 13 follows from a theorem due to Bôcher and Walsh([46]).

THEOREM 14. *Let the function $w=f(z)$ analytic for $|z|<r$ with $f(0)=0$ map $|z|<r$ onto a Riemann configuration $R$ such that no point of the boundary of $R$*

([45]) The part of this theorem which refers to the zeros of $f'(z)$ is not new, if $q$ is defined as the number of zeros of $f(z)$ in $B$, and has been considered by de Boer, Macdonald, de la Vallée Poussin, Whittaker and Watson, Denjoy, Lange-Nielsen, and Ålander. See for instance Denjoy, Comptes Rendus de l'Académie des Sciences, Paris, vol. 166 (1918), pp. 31–33; Ålander, Comptes Rendus de l'Académie des Sciences, Paris, vol. 184 (1927), pp. 1411–1413.

([46]) J. L. Walsh, these Transactions, vol. 19 (1918), pp. 291–298, especially p. 297.

*satisfies the inequality* $|w| < \rho > 0$. *Then the connected region* $R_1$ *of* $R$ *which contains the transform of* $z=0$ *and which is cut out of* $R$ *by a biscuit-cutter whose center is* $w=0$ *and radius* $\rho$ *is simply connected; and each point* $w$ *of the* $w$*-plane with* $|w| < \rho$ *is covered by* $R_1$ *the same number of times.*

The region $R_1$ corresponds to some region $R_2$ in $|z| < r$ containing $z=0$. The function $|f(z)|$ is continuous in the closed region consisting of $R_2$ plus its boundary, and assumes the constant value $\rho$ on the boundary; of course the boundary of $R_2$ may coincide in whole or in part with $|z| = r$. It follows from the principle of maximum modulus applied to $f(z)$ in $|z| < r$ that the boundary of $R_2$ cannot fall into two or more continua, one of which would necessarily lie in a simply connected region interior to $|z| = r$ bounded by another continuum belonging to the boundary of $R_2$. Then $R_2$ is simply connected, and so consequently is $R_1$. The remainder of Theorem 14 follows from Theorem 13.

THEOREM 15. *Let* $w=f(z)$ *be analytic for* $|z| < 1$ *and map* $|z| < 1$ *onto the Riemann configuration* $R$ *with* $w_0=f(z_0)$, $|z_0| < 1$. *Let* $f(z)$ *take on in* $|z| < 1$ *every value* $w$ *in the region* $|w-w_0| < \rho > 0$ *precisely* $p$ *times. Then we have* $D_p(w_0) \geqq \rho$.

No boundary point $w_1$ of $R$ can satisfy the inequality $|w_1-w_0| < \rho$; for if it did the point $w_1$ of the $w$-plane would be covered by $R$ a totality of $p$ times, and by the implicit function theorem a suitably chosen neighborhood of $w_1$ would also be covered by $R$ precisely $p$ times by the sheets of $R$ covering $w_1$. Some values $w$ in every neighborhood of $w_1$ are covered also by the sheet (or sheets) of $R$ of which $w_1$ is a boundary point; so some points $w$ with $|w-w_0| < \rho$ are covered more than $p$ times, contrary to hypothesis.

We have now shown that no boundary point of $R$ satisfies the inequality $|w-w_0| < \rho$; so it follows from Theorem 14 that the region containing $w_0$ cut out of $R$ by a biscuit-cutter of center $w_0$ and radius $\rho$ covers each point of $|w-w_0| < \rho$ the same number of times, a number which by the hypothesis of Theorem 15 cannot exceed $p$; hence Theorem 15 is established.

Still another result related to Theorems 14 and 15 follows easily:

THEOREM 16. *Let the function* $w=f(z)$ *be analytic for* $|z| < 1$ *and map* $|z| < 1$ *onto the Riemann configuration* $R$ *with* $w_0=f(z_0)$, $|z_0| < 1$. *Suppose* $\liminf_{|z| \to 1} |f(z)-w_0| \geqq \rho$, *and suppose no value in* $|w-w_0| < \rho$ *is taken on by* $f(z)$ *in* $|z| < 1$ *more than* $p$ *times. Then we have* $D_p(w_0) \geqq \rho$.

It follows from our hypothesis that no boundary point of $R$ lies in $|w-w_0| < \rho$; so Theorem 16 follows from Theorems 14 and 15.

COROLLARY. *Let* $w=f(z)$ *be analytic for* $|z| < 1$ *and map* $|z| < 1$ *onto the Riemann configuration* $R$ *with* $w_0=f(z_0)$, $|z_0| < 1$. *Let* $R_1$ *be a subregion of* $|z| < 1$ *containing* $z_0$, *whose boundary* $B$ *satisfies the condition* $\lim_{z \to B, |z| < 1} |f(z)-w_0|$

$=\rho>0$, *and suppose no value* $w$ *is taken on by* $f(z)$ *in* $R_1$ *more than* $p$ *times. Then we have* $D_p(w_0)\geqq\rho$.

It follows from the principle of maximum modulus that $R_1$ is simply connected. If $R_1$ is mapped smoothly and conformally onto $|\zeta|<1$, and if Theorem 16 is applied to the function which maps $|\zeta|<1$ onto $R_1$, we obtain the corollary.

Although the following theorem is not needed in the sequel, it is of some interest in itself.

THEOREM 17. *Let* $R$ *be a simply connected Riemann configuration of hyperbolic type, and let* $w_0$ *be any point of* $R$. *Then* $D_p(w_0)$ *is a continuous function of* $w_0$.

We need to define what we shall mean by the continuity of $D_p(w_0)$ on $R$. If $w_0$ is a branch point of order greater than $p-1$, $(p>1)$, we shall say that $D_p(w_0)$ is continuous at $w_0$ if to any $\epsilon>0$ we can assign a number $\delta>0$ so that for any point $w_0'$ at a distance not greater than $\delta$ from $w_0$ and lying on one of the sheets that come together at $w_0$ the relation $|D_p(w_0')-D_p(w_0)|<\epsilon$ holds; here $D_p(w_0)=0$. If $w_0$ is a branch point of order $q$, where $0\leqq q\leqq p-1$, we shall say that $D_p(w_0)$ is continuous at $w_0$ if to any $\epsilon>0$ we can assign a number $\delta>0$ so that for any point $w_0'$ within the $q$-sheeted circle $C_q$ with center at $w_0$ and radius $\delta$ the relation $|D_p(w_0')-D_p(w_0)|<\epsilon$ holds. The proof of this theorem is left to the reader.

**16. The limit property of $D_p$ for continuously convergent sequences.**

THEOREM 18. *Let* $\{f_n(z)\}$ *be a sequence of functions analytic in the unit circle* $|z|<1$, *and converging uniformly in every closed subregion of* $|z|<1$ *to an analytic function* $f(z)$. *Let* $z_0$ *be any point in the circle* $|z|<1$ *and set* $w_n=f_n(z_0)$, $w_0=f(z_0)$. *Denoting by* $D_p(w_n)$ *the radius of* $p$*-valence at the point* $w_n$ *of the Riemann configuration* $R_n$ *on which* $f_n(z)$ *maps the circle* $|z|<1$ *and by* $D_p(w_0)$ *the radius of* $p$*-valence at the point* $w_0$ *of the Riemann configuration* $R_0$ *on which* $f(z)$ *maps the circle* $|z|<1$, *we have*

$$\lim_{n\to\infty} D_p(w_n) = D_p(w_0), \qquad p = 1, 2, 3, \cdots. \tag{16.1}$$

The proof of the theorem will be based on two lemmas:

LEMMA 1. *Under the conditions of Theorem* 18,

$$\liminf_{n\to\infty} D_p(w_n) \geqq D_p(w_0). \tag{16.2}$$

The lemma is clearly trivial if $D_p(w_0)=0$.

Let us assume, therefore, that $D_p(w_0)>0$ and choose any positive number $\rho$ so that $\rho<D_p(w_0)$. Hence, the Riemann configuration $R_0$ contains in its interior some $q$-sheeted circle $C_q(w_0)$, $1\leqq q\leqq p$, of center $w_0$ and radius $\rho$, to-

gether with its boundary. Denote the region in $|z|<1$ on which the function $w=f(z)$ maps $C_q(w_0)$ by $R_0$. The boundary $B_0$ of $R_0$ must consequently lie wholly in the interior of $|z|<1$. In $R_0$ we have $|f(z)-w_0|<\rho$ and on $B_0$ we have $|f(z)-w_0|=\rho$. Let $\epsilon>0$ be any number such that $\rho+\epsilon<D_p(w_0)$. Due to the uniform convergence of the sequence $f_n(z)$ on $B_0$, there exists a positive integer $n(\epsilon)$ such that for all integers $n>n(\epsilon)$ the inequality $|f_n(z)-w_n|>\rho-\epsilon$ holds on $B_0$. Hence, that region $R_n$ in the circle $|z|<1$ which contains the point $z_0$ and on which $|f_n(z)-w_n|<\rho-\epsilon$ lies wholly interior to $R_0$. On the boundary $B_n$ of $R_n$ we have $|f_n(z)-w_n|=\rho-\epsilon$. In accordance with Theorem 13 the function $f_n(z)$ takes on all its values the same number of times $q_n$ in $R_n$. By Hurwitz's theorem, since $f(z)$ is at most $p$-valent([47]) in $R_0$, for sufficiently large values of $n$ we have $q_n\leqq p$. Hence, by the corollary to Theorem 16 we have $D_p(w_n)\geqq D_{q_n}(w_n)\geqq\rho-\epsilon$. Hence, $\liminf_{n\to\infty} D_p(w_n)\geqq\rho$. But $\rho$ is an arbitrary positive number less than $D_p(w_0)$. The relation (16.2) follows at once.

LEMMA 2. *Under the conditions of Theorem* 18,

$$\limsup_{n\to\infty} D_p(w_n) \leqq D_p(w_0). \tag{16.3}$$

If this lemma is false there must exist a positive constant $a$ such that for infinitely many values of $n$

$$D_p(w_n) > a > D_p(w_0). \tag{16.4}$$

We shall neglect all those functions $f_n(z)$ for which the above inequality fails and assume that (16.4) holds for all $n$.

Consider that largest region $R_n$ in the circle $|z|<1$ which contains the point $z_0$, for which $|f_n(z)-w_n|<a$. Then in $R_n$ the function $f_n(z)$ is $q$-valent $(q\leqq p)$. According to (16.4) the boundary $C_n$ of the region $R_n$ lies wholly in the circle $|z|<1$. Furthermore, by the principle of maximum modulus we conclude that $R_n$ is simply connected. Clearly, on the curve $C_n$ the relation $|f_n(z)-w_n|=a$ is satisfied. Every value taken on by $f_n(z)$ in $R_n$ is taken on the same number of times.

Denote by $z=\phi_n(t)$ a function which maps the region $R_n$ on the circle $|t|<1$ in such a manner that $\phi_n(0)=z_0$. Since the curve $C_n$ is a Jordan curve, by a well known theorem of Osgood-Carathéodory the function $\phi_n(t)$ is continuous in the closed circle $|t|\leqq 1$([48]). The function $f_n(\phi_n(t))=g_n(t)$ is analytic in $|t|<1$, continuous in $|t|\leqq 1$ and $|g_n(t)-w_n|=a$ on $|t|=1$. By Schwarz's reflection principle([49]), we infer that $g_n(t)$ is analytic in the closed circle

([47]) We shall say that a function $f(z)$ is *$p$-valent in a region* $R$ if it assumes no value more than $p$ times in $R$ and at least one value precisely $p$ times. A function $f(z)$ will be called *at most $p$-valent in* $R$ if it is $q$-valent in $R$ for some $q\leqq p$.

([48]) W. F. Osgood and E. H. Taylor, these Transactions, vol. 14 (1913), pp. 277–298; C. Carathéodory, Mathematische Annalen, vol. 73 (1913), pp. 305–320.

([49]) Cf. G. Julia, loc. cit., p. 44 ff.

$|t| \leqq 1$. Finally, $g_n(t)$ is precisely $q$-valent in $|t| < 1$ since $f_n(z)$ possesses the same property in $R_n$. By the theorem of Radó, referred to earlier, we may represent $g_n(t)$ in the following manner:

$$g_n(t) = w_n + ae^{i\theta_n}t^{k_n} \prod_{j=1}^{q-k_n} \frac{t - t_j^{(n)}}{1 - \overline{t_j^{(n)}}t}, \quad k_n \geqq 1; \ |t_j^{(n)}| \leqq 1. \tag{16.5}$$

Since the $g_n(t)$ are uniformly bounded, they form a normal family and we may select a subsequence, which for simplicity will again be denoted by $\{g_n(t)\}$, converging uniformly in every closed subregion of $|t| < 1$ to a function $G(t)$ analytic in $|t| < 1$. On account of (16.5) $G(t)$ has itself a representation of the form

$$G(t) = w_0 + ae^{i\theta_0}t^k \prod_{j=1}^{q-k} \frac{t - t_j}{1 - \bar{t}_j t}, \qquad k \geqq 1. \tag{16.6}$$

Just as in (16.5) some of the $t_j$ here may have the absolute value 1.

Now consider that largest region $R_0$ in $|z| < 1$ which contains the point $z_0$ and in which $|f(z) - w_0| < a$. According to the maximum modulus principle $R_0$ is simply connected and we may map it on the circle $|t| < 1$ by means of a function $z = \phi_0(t)$ so that $\phi_0(0) = z_0$. On that part of the boundary $B_0$ of $R_0$ which lies interior to the circle $|z| < 1$ if it exists we have $|f(z) - w_0| = a$. We shall now show that $R_0$ is the kernel of the sequence of regions $\{R_n\}$ (50). Indeed, consider any region $R_0'$ which together with its boundary lies interior to $R_0$ and contains the point $z_0$. By the definition of $R_0$, in the region $R_0'$ and on its boundary we have $|f(z) - w_0| < a$. Since the functions $f_n(z) - w_n$ converge uniformly to $f(z) - w_0$ in the closure of $R_0'$, for $n$ sufficiently large we have $|f_n(z) - w_n| < a$ in the closure of $R_0'$, and therefore $R_0'$ belongs to all $R_n$ for sufficiently large values of $n$. Next, choose any point $z'$ of the circle $|z| < 1$ exterior to $R_0$ (if such a point exists). Connect the point $z'$ with the point $z_0$ by any Jordan arc $L$ which lies wholly in the circle $|z| < 1$. Since $z'$ is exterior to $R_0$, there must exist on the arc $L$ at least one point $Z$ at which $|f(Z) - w_0| > a$. For sufficiently large values of $n$ we must have $|f_n(Z) - w_n| > a$, and consequently $Z$ is exterior to $R_n$. Thus, on any Jordan arc joining the points $z_0$ and $z'$ there exists a point exterior to $R_n$ for all sufficiently large values of $n$. Consequently $R_0$ is the kernel of the sequence of regions $\{R_n\}$. Hence, by a well known theorem of Carathéodory(51) the sequence of functions $\phi_n(t)$ converges uniformly in every closed subregion of $|t| < 1$ to the function $\phi_0(t)$, provided merely we have chosen $\phi_n'(0) > 0$, $\phi_0'(0) > 0$.

If we form the function $g_0(t) = f(\phi_0(t))$, it follows that the sequence of func-

(50) For the notion of kernel of a sequence of domains cf. C. Carathéodory, *Conformal Representation*, Cambridge Tract in Mathematics and Mathematical Physics, no. 28, (1932), pp. 74–77.

(51) C. Carathéodory, loc. cit., particularly p. 76.

tions $\{g_n(t)\}$ converges uniformly in every closed subregion of $|t|<1$ to the function $g_0(t)$. We have shown earlier, however, that the sequence $\{g_n(t)\}$ converges to the function $G(t)$ whose representation is given in (16.6). We thus find that $g_0(t)=G(t)$ identically in $|t|<1$. From (16.6) it follows therefore that $g_0(t)$ is analytic in $|t|\leqq 1$, is $q'$-valent ($q'\leqq p$) in $|t|<1$, and on the circumference $|t|=1$ satisfies the relation $|g_0(t)-w_0|=a$.

Consider now an arbitrary positive number $\epsilon$ such that $a-\epsilon>D_p(w_0)$. Denote by $R_\epsilon$ the largest region in $|t|<1$ which contains the origin and throughout which $|g_0(t)-w_0|<a-\epsilon$. The boundary $C_\epsilon$ of this region lies wholly interior to $|t|<1$ and in $R_\epsilon$ the function $g_0(t)$ is $q''$-valent ($q''\leqq p$). The function $z=\phi_0(t)$ maps the region $R_\epsilon$ on a region $P_\epsilon$ in the $z$-plane which is together with its boundary $\Gamma_\epsilon$ interior to $R_0$. In $P_\epsilon$ we have $|f(z)-w_0|<a-\epsilon$ and on $\Gamma_\epsilon$ we have $|f(z)-w_0|=a-\epsilon$. Since the region $P_\epsilon$ contains the point $z_0$ and since $f(z)$ is $q''$-valent in $P_\epsilon$, it follows that $a-\epsilon<D_p(w_0)$. This contradicts our assumption concerning $\epsilon$.

Since the assumption (16.4) leads to a contradiction, the relation (16.3) is true.

We are now ready to prove the theorem. Lemmas 1 and 2 together yield the inequalities

$$\limsup_{n\to\infty} D_p(w_n) \leqq D_p(w_0) \leqq \liminf_{n\to\infty} D_p(w_n).$$

Since, however, we always have $\lim\inf_{n\to\infty} D_p(w_n)\leqq\lim\sup_{n\to\infty} D_p(w_n)$, it follows that $\lim\sup_{n\to\infty} D_p(w_n)=\lim\inf_{n\to\infty} D_p(w_n)=\lim_{n\to\infty} D_p(w_n)=D_p(w_0)$, which proves the theorem.

17. **$\lim_{n\to\infty} D_p(w_n)=0$ is a necessary and sufficient condition for $\lim_{n\to\infty}|f^{(k)}(z_n)|(1-|z_n|)^k=0$ $(k=1, 2, \cdots, p)$.** An immediate consequence of Theorem 18 is the following extension of Theorem 2, Chapter II, to the higher derivatives of bounded functions.

THEOREM 19. *Let $f(z)$ be regular and bounded in $|z|<1$:*

$$|f(z)| \leqq M,$$

*let $\{z_n\}$ be any sequence of points in $|z|<1$, and let $w_n=f(z_n)$. Then, a necessary and sufficient condition for*

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \qquad k = 1, 2, \cdots, p,$$

*is that* $\lim_{n\to\infty} D_p(w_n)=0$.

We first prove the sufficiency of the condition. We assume that $\lim_{n\to\infty} D_p(w_n)=0$. In accordance with the definition of the radius of $p$-valence it follows that

$$\lim_{n\to\infty} D_k(w_n) = 0, \qquad k = 1, 2, \cdots, p. \tag{17.1}$$

By virtue of Theorem 2, Chapter II, the condition is sufficient for $p=1$. Let us assume that the condition is sufficient for $p-1$ and prove it to be sufficient for $p$. We assume therefore that (17.1) implies

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \quad k = 1, 2, \cdots, p-1. \tag{17.2}$$

If the condition is not sufficient for $p$, we could find a positive constant $\delta$ and a subsequence of $\{z_n\}$, which for simplicity will again be denoted by $\{z_n\}$ for which

$$|f^{(p)}(z_n)|(1-|z_n|)^p \geqq \delta > 0 \tag{17.3}$$

and at the same time the relations (17.1) and (17.2) hold.

Now if we introduce the sequence of functions

$$\phi_n(\zeta) = f\left(\frac{\zeta + z_n}{1 + \bar{z}_n\zeta}\right)$$

which are bounded and regular in $|\zeta| < 1$: $|\phi_n(\zeta)| \leqq M$, we obtain by virtue of the expression (2.3)

$$\frac{\phi_n^{(p)}(0)}{p!} = \sum_{\nu=0}^{p-1} (-1)^\nu C_{p-1,\nu}\bar{z}_n^\nu \frac{(1-|z_n|^2)^{p-\nu} f^{(p-\nu)}(z_n)}{(p-\nu)!}.$$

The relations (17.2) and (17.3) imply

$$\liminf_{n\to\infty} |\phi_n^{(p)}(0)| \geqq \delta > 0, \tag{17.4}$$

while the relation (2.3) written out for $n=1, 2, \cdots, p-1$ together with (17.2) shows that

$$\lim_{n\to\infty} |\phi_n^{(k)}(0)| = 0, \qquad \text{for } k = 1, 2, \cdots, p-1. \tag{17.5}$$

The sequence of functions $\{\phi_n(\zeta)\}$ forms a normal family in $|\zeta| < 1$. We may, therefore, extract a convergent subsequence which for simplicity will again be denoted by $\{\phi_n(\zeta)\}$

$$\lim_{n\to\infty} \phi_n(\zeta) = \phi(\zeta).$$

The relations (17.4) and (17.5) imply

$$\phi^{(k)}(0) = 0 \quad \text{for} \quad k = 1, 2, \cdots, p-1; \qquad |\phi^{(p)}(0)| \geqq \delta. \tag{17.6}$$

The equations in (17.6), however, imply that the radius of $p$-valence $D_p[\phi(0)]$ of the Riemann surface on which $\phi(\zeta)$ maps the circle $|\zeta| < 1$ is positive at the point $\phi(0)$ of the surface $D_p[\phi(0)] > 0$. According to Theorem 18 if we

denote by $D_p[\phi_n(0)]$ the radius of $p$-valence at the point $\phi_n(0)$ of the Riemann surface $R_n$ on which $\phi_n(\zeta)$ maps the circle $|\zeta|<1$ and observe that $\phi_n(0)=w_n$, we obtain

$$\lim_{n\to\infty} D_p(w_n) = D_p[\phi(0)] > 0.$$

But $R_n$ is precisely the Riemann surface $R$ on which $f(z)$ maps the circle $|z|<1$. Hence, the last relation contradicts (17.1) for $k=p$. This proves that the assumption (17.3) is false and the sufficiency of our condition is established.

We now turn to the proof of the necessity of the condition in Theorem 19. Let us assume that

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \quad \text{for } k = 1, 2, \cdots, p.$$

Forming again the functions $\phi_n(\zeta)$, we see that

$$\lim_{n\to\infty} |\phi_n^{(k)}(0)| = 0 \qquad \text{for } k = 1, 2, \cdots, p.$$

Let us assume that we have already selected a uniformly convergent subsequence of the $\{\phi_n(\zeta)\}$, which, because of the normality of the family, is always possible. The limit function $\phi(\zeta)$ of the sequence has the property that $\phi^{(k)}(0) = 0$ for $k = 1, 2, \cdots p$. Consequently, $D_p[\phi(0)]=0$ and by Theorem 18

$$\lim_{n\to\infty} D_p(w_n) = 0.$$

The last relation has been proved only for a subsequence of the original sequence. But since from every sequence we may select a subsequence with this property, it must also hold for the whole sequence. Theorem 19 is now established.

It will be noticed that Theorem 19 is unsatisfactory in that no indication is given of the manner in which expressions of the type $|f^{(k)}(z_n)|(1-|z_n|)^k$ depend on the radii of $p$-valence $D_p(w_n)$. In the case $p=1$ we have already given inequalities which bring out this dependence (Theorem 3, Chapter II). Our next task will be to extend Theorem 3, Chapter II, to the higher derivatives of bounded functions. The constants that we shall obtain will, however, not be precise. We shall first study upper bounds for the derivatives of bounded functions. The inequalities that we shall obtain will, of course, yield a new proof of Theorem 19 by quantitative methods rather than the purely qualitative methods that we used in the present proof.

## CHAPTER III. BOUNDED FUNCTIONS; INEQUALITIES ON $D_p$

18. **A preliminary lower bound for $D_p$.** For our purpose in the use of $D_p(w_0)$ for the study of such relations as $|f^{(p)}(z_k)|(1-|z_k|)^p\to 0$, it is desir-

able to have explicit numerical inequalities connecting $D_p(w_k)$ and the derivatives $f'(z_k), f''(z_k), \cdots, f^{(p)}(z_k)$. We first prove regarding this relationship

THEOREM 1. *Suppose the function $f(z)$ analytic for $|z|<1$ with $f(0)=0$, $f^{(p)}(0)=p!$, and with $|f(z)| \leqq M$ for $|z|<1$. Then we have*

$$D_p(0) \geqq M_p > 0, \tag{18.1}$$

*where $M_p = M_p(M)$ is a suitably chosen constant depending on $M$ and $p$ but not on $f(z)$.*

Our proof of Theorem 1 is a direct generalization of Landau's proof([82]) for the case $p=1$. For the case $p=1$, Landau's method yields the inequality

$$D_1(0) \geqq 1/(6M), \tag{18.2}$$

a special case of inequality (18.9) to be proved below. But other related methods([83]) yield the inequality

$$D_1(0) \geqq 1/(4M), \tag{18.3}$$

which is somewhat sharper than (18.2) and which we shall therefore take as point of departure.

We remark that if $f(z)$ is analytic for $|z|<1$ with $f(0)=0$, $f'(0)=m\neq 0$, with $|f(z)| \leqq M$ for $|z|<1$, then the function $f(z)/m$ has the derivative unity at the origin and modulus in $|z|<1$ not greater than $M/|m|$. Consequently under the transformation $w=f(z)/m$ we have from (18.3) the result $D_1(0) \geqq |m|/(4M)$, and under the transformation $w=f(z)$ we have

$$D_1(0) \geqq \frac{|m^2|}{4M}. \tag{18.4}$$

Let us now suppose Theorem 1 established with $p$ replaced by $j$ for $j=1, 2, \cdots, p-1$; we proceed to prove by induction the theorem as stated. The cases

$$|f'(0)| \geqq \frac{1}{(12M)^{p-1}}, \tag{18.5'}$$

$$\frac{|f''(0)|}{2!} \geqq \frac{1}{(12M)^{p-2}}, \tag{18.5''}$$

$$\cdots\cdots\cdots\cdots,$$

$$\frac{|f^{(p-1)}(0)|}{(p-1)!} \geqq \frac{1}{12M} \tag{18.5$^{(p-1)}$}$$

are all handled in a manner similar to the proof of (18.4). Thus in case

---

([82]) E. Landau, Sitzungsberichte der Königlichen Preussischen Akademie der Wissenschaften, Berlin, 1926, pp. 467–474.

([83]) J. Dieudonné, Annales de l'École Normale Supérieure, vol. 48 (1931), pp. 247–358. Or see Montel, *Fonctions Univalentes*, §37.

$(18.5^{(j)}), j=1, 2, \cdots, p-1$, the function

$$\frac{f(z)}{f^{(j)}(0)/j!} \tag{18.6}$$

has the $j$th derivative $j!$ at the origin, and modulus in $|z|<1$ not greater than $j!M/|f^{(j)}(0)|$, so by our assumption that Theorem 1 with $p$ replaced by $j$ is established, we have under the transformation $w=j!f(z)/f^{(j)}(0)$, $D_j(0) \geqq M_j(j!M/|f^{(j)}(0)|)$, and we have under the transformation $w=f(z)$

$$D_j(0) \geqq \frac{|f^{(j)}(0)|}{j!}\cdot M_j\left[\frac{j!M}{|f^{(j)}(0)|}\right]; \tag{18.7}$$

hence by the relation $D_p(0) \geqq D_j(0)$ the theorem may be considered to be proved. It remains to study the case that we have simultaneously

$$\frac{|f^{(j)}(0)|}{j!} < \frac{1}{(12M)^{p-j}}, \qquad j = 1, 2, \cdots, p-1, \tag{18.5^{(p)}}$$

with, of course, the relation $f^{(p)}(0)=p!$.

Suppose $r$ can be chosen $(0<r<1)$ so that the expression

$$R = r^p - \max_{|z|=r} |f(z) - z^p| \tag{18.8}$$

is positive. Then we have $R \leqq r^p < r$, and for $|w|<R$ the inequality

$$\left|\frac{f(z) - z^p}{z^p - w}\right| < 1$$

holds on the circle $|z|=r$. Of course, $z^p-w$ cannot vanish on $|z|=r$. Then by Rouché's theorem the function $f(z)-w$ has precisely as many zeros in $|z|<r$ as does the function $z^p-w$, namely $p$. Then the transformation $w=f(z)$ maps $|z|<r$ onto a Riemann configuration which contains the region $|w|<R$ with each point covered precisely $p$ times. Thus (Theorem 15, Chapter II), we have $D_p(0) \geqq R$, whether $D_p(0)$ refers to the Riemann configuration which is the image of $|z|<r$ or to the configuration which is the image of $|z|<1$ under the transformation $w=f(z)$.

It remains to show that $r$ can be chosen in such a way that $R$ as defined by (18.8) is positive. If we set $f(z) \equiv \sum_{n=1}^{\infty} a_n z^n$, Cauchy's inequality is $|a_n| \leqq M$, and in particular $a_p = 1 \leqq M$. Consequently we may write on $|z|=r$ by the use of $(18.5^{(p)})$

$$\left|\sum_{n=p+1}^{\infty} a_n z^n\right| \leqq \frac{Mr^{p+1}}{1-r},$$

$$\left|\sum_{n=1}^{p-1} a_n z^n\right| \leqq \sum_{n=1}^{p-1} \frac{r^n}{(12M)^{p-n}} = \frac{r}{(12M)^{p-1}}\,\frac{1-(12M)^{p-1}r^{p-1}}{1-12Mr},$$

and with the choice $r=1/(4M)$,

$$
\begin{aligned}
R &= r^p - \max_{|z|=r} \left| f(z) - z^p \right| \\
&\geqq r^p - \frac{Mr^{p+1}}{1-r} - \frac{r}{(12M)^{p-1}} \, \frac{1-(12M)^{p-1}r^{p-1}}{1-12Mr} \\
&\geqq r^p - \frac{4}{3} Mr^{p+1} - \frac{r}{(12M)^{p-1}} \, \frac{1-(12M)^{p-1}r^{p-1}}{1-12Mr} \qquad (18.9)\\
&= \frac{1+3^{p-2}}{2\cdot 3^{p-1}\cdot 4^p \cdot M^p}\,.
\end{aligned}
$$

We have now proved the desired inequality $R>0$ and thus completed the proof of Theorem 1, and we also have material for obtaining an explicit inequality for $M_p(M)$ in inequality (18.1).

19. **Numerical lower bounds for $D_p$.** When $p=2$, relation (18.7) [or (18.4)] becomes in case (18.5′)

$$D_1(0) \geqq \frac{1}{24^2M^3},$$

whereas in case (18.5″) we have from (18.9)

$$D_2(0) \geqq \frac{1}{48M^2},$$

so in either case we may write

$$D_2(0) \geqq \frac{1}{24^2M^3}\,. \tag{19.1}$$

Inequality (19.1) is to be generalized by proving

$$D_p(0) \geqq M_p(M) \equiv \frac{1}{4\cdot 12^{2^p-2}M^{2^p-1}}\,. \tag{19.2}$$

We remark that $M_p(M)$, as thus defined, decreases monotonically as $M$ increases. It is to be noticed that (19.2) holds for $p=1$, by inequality (18.3), and for $p=2$, by inequality (19.1); we assume (19.2) to hold with $p$ replaced by $j$ for $j=2, 3, \cdots, p-1$, and shall establish (19.2) as written. In case $(18.5^{(j)})$ we find from (18.7) and (19.2) the inequality ($p>2$)

$$D_j(0) \geqq \frac{1}{(12M)^{p-j}} \, \frac{1}{4\cdot 12^{2^j-2}[M(12M)^{p-j}]^{2^j-1}}\,. \tag{19.3}$$

Direct comparison of the right-hand members of (19.2) and (19.3) now shows,

by virtue of the inequality $2^{q-1} \geqq q$, $q$ a positive integer, and by virtue of $D_p(0) \geqq D_j(0)$, that (19.2) holds in each of the cases $(18.5^{(j)})$, $j=1, 2, \cdots, p-1$. Also in case $(18.5^{(p)})$ inequality (19.2) is valid, as we find from (18.9), so we have established.

COROLLARY 1. *Under the hypothesis of Theorem* 1, *we have inequality* (19.2).

Needless to say, the numerical results contained in some of the preceding inequalities can be improved, and it is to be supposed that those contained in inequality (19.2) can be greatly improved.

Inequality (18.7) is valid under the assumption $f^{(j)}(0) \neq 0$ instead of $f^{(j)}(0) = j!$, so by using $M_p(M)$ as defined by (19.2) we may formulate:

COROLLARY 2. *Suppose the functions $f_k(z)$ analytic for $|z| < 1$, with $f_k(0) = 0$ and $|f_k(z)| \leqq M$ for $|z| < 1$. If as $k$ becomes infinite the corresponding sequence $D_p(0)$ approaches zero, then we have also*

$$\lim_{k\to\infty} f_k^{(p)}(0) = 0.$$

Under the conditions of Corollary 2 we have $D_p(0) \geqq D_{p-1}(0) \geqq \cdots \geqq D_1(0)$, from which follows for $j = 1, 2, \cdots, p$ the relation

$$\lim_{k\to\infty} f_k^{(j)}(0) = 0. \tag{19.4}$$

A specific inequality for the direct proof of (19.4) is useful. A consequence of (19.2) and (18.7) for $j = 1, 2, \cdots$, p, with the omission of the requirement $f^{(j)}(0) = j!$, is

$$D_j(0) \geqq \frac{|f^{(j)}(0)|}{j!} \frac{1}{4 \cdot 12^{2^j-2}(j!M/|f^{(j)}(0)|)^{2^j-1}}.$$

The inequality $D_j(0) \leqq M$ is obvious, so we have

$$\begin{aligned} \frac{|f^{(j)}(0)|}{j!} &\leqq 4^{2^{-j}} \cdot 12^{1-2^{1-j}} M^{1-2^{-j}} [D_j(0)]^{2^{-j}} \\ &\leqq 24M \left[\frac{D_j(0)}{M}\right]^{2^{-j}} \\ &\leqq 24M \left[\frac{D_j(0)}{M}\right]^{2^{-p}}. \end{aligned} \tag{19.5'}$$

By virtue of the inequalities $D_j(0) \geqq D_{j-1}(0)$ we may now write[84]

$$|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)| \leqq 24p(1+M)[D_p(0)]^{2^{-p}}. \tag{19.5}$$

[84] For $0 < \alpha \leqq 1$, $M > 0$, we have $M^\alpha \leqq 1 + M$.

We state explicitly a major result:

COROLLARY 3. *If $f(z)$ is analytic and in modulus not greater than $M$ for $|z|<1$, with $f(0)=0$, then inequality* (19.5) *is valid for every positive integer $p$.*

For the purpose of Corollary 3, the factor $1+M$ in the right-hand member of (19.5) may of course be replaced by $M^{1-2^{-p}}$.

20. **A lower bound for the derivative of a circular product.** The converse of Corollary 2 is false, as is illustrated by the sequence $f_k(z)\equiv z$, with $p=2$. The second derivative $f_k''(0)$ vanishes for every $k$, yet $D_2(0)$ has the constant value unity, so the relation $D_2(0)\to 0$ is not satisfied. Indeed, in the general situation that $f_k(z)$ is analytic for $|z|<1$ with $f_k(0)=0$ and $|f_k(z)|\leqq M$ for $|z|<1$, it is not to be expected that $f_k^{(p)}(0)\to 0$ should imply $D_p(0)\to 0$, for the latter relation by virtue of $D_l(0)\geqq D_{l-1}(0)$ implies also $D_l(0)\to 0$, $l=1, 2, \cdots, p-1$ which by Corollary 2 implies (19.4), a relation which is not implied by the hypothesis and is indeed completely independent of the hypothesis. We should expect, then, that a relation in the opposite sense to Corollary 2 would necessarily involve the lower derivatives. We shall proceed to prove

THEOREM 2. *Let the function $w=f(z)$ analytic for $|z|<1$ map $|z|<1$ onto a Riemann configuration with $f(0)=0$. Then there exists a positive constant $\gamma_p$ depending on $p$ but not on $f(z)$ such that we have*

$$(20.1)\qquad D_p(0)\leqq\frac{1}{\gamma_p}\left[|f'(0)|+\frac{1}{2!}|f''(0)|+\cdots+\frac{1}{p!}|f^{(p)}(0)|\right].$$

The proof of Theorem 2 is to be carried out in several steps, of which the first is

THEOREM 3. *Let $w=g(z)$ analytic for $|z|<1$ map $|z|<1$ onto $|w|<1$ counted precisely $p$ times, or precisely $m<p$ times, with $g(0)=0$. Then we have*

$$(20.2)\qquad |g'(0)|+\frac{1}{2!}|g''(0)|+\cdots+\frac{1}{p!}|g^{(p)}(0)|\geqq c_p>0,$$

*where $c_p$ is a suitably chosen number depending on $p$ but not on $g(z)$.*

To be explicit, we prove (20.2) with $c_p=2^{-(p+1)!}$.

The most general function $g(z)$ is of the form[55]

$$(20.3)\qquad w=g_p(z)=z\prod_{j=1}^{p-1}\frac{z-\beta_j}{1-\bar{\beta}_jz},\qquad |\beta_j|\leqq 1,$$

except for a constant factor of modulus unity which does not affect the left-hand member of (20.2) and which we therefore suppress. In the case $p=1$, the

[55] T. Radó, ibid.

form (20.3) breaks down, but we have $g_1(z)\equiv z$, and (20.2) is fulfilled with $c_p$ replaced by unity, which is greater than $c_1=1/4$. Henceforth, we suppose $p\geqq 2$.

We prove (20.2) by induction, assuming the validity of (20.2) with $p$ replaced by $p-1$ and proving (20.2) as written(56). Equation (20.3) can be expressed in the equivalent form

$$g_p(z) = g_{p-1}(z)\,\frac{z-\alpha}{1-\bar{\alpha}z}, \qquad |\alpha| \leqq 1, \tag{20.4}$$

where we have also

$$g_{p-1}(z) = a_1 z + a_2 z^2 + \cdots, \qquad |z| < 1,$$
$$g_p(z) = b_1 z + b_2 z^2 + \cdots, \qquad |z| < 1.$$

The power series expansions of the second factor in the right-hand member of (20.4) and of its reciprocal yield by direct comparison of coefficients the two sets of equations

$$\begin{aligned}
b_1 &= -a_1\alpha,\\
b_2 &= a_1(1-\alpha\bar{\alpha}) - a_2\alpha,\\
b_3 &= a_1\bar{\alpha}(1-\alpha\bar{\alpha}) + a_2(1-\alpha\bar{\alpha}) - a_3\alpha,\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
b_k &= a_1\bar{\alpha}^{k-2}(1-\alpha\bar{\alpha}) + a_2\bar{\alpha}^{k-3}(1-\alpha\bar{\alpha}) + \cdots + a_{k-1}(1-\alpha\bar{\alpha}) - a_k\alpha;
\end{aligned} \tag{20.5}$$

$$\begin{aligned}
a_1 &= -\frac{b_1}{\alpha}, \qquad\qquad \alpha \neq 0,\\
a_2 &= -b_1\frac{1-\alpha\bar{\alpha}}{\alpha^2} - \frac{b_2}{\alpha},\\
a_3 &= -b_1\frac{1-\alpha\bar{\alpha}}{\alpha^3} - b_2\frac{1-\alpha\bar{\alpha}}{\alpha^2} - \frac{b_3}{\alpha},\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,\\
a_k &= -b_1\frac{1-\alpha\bar{\alpha}}{\alpha^k} - b_2\frac{1-\alpha\bar{\alpha}}{\alpha^{k-1}} - \cdots - b_{k-1}\frac{1-\alpha\bar{\alpha}}{\alpha^2} - \frac{b_k}{\alpha}.
\end{aligned} \tag{20.6}$$

---

(56) The succeeding proof can be considerably shortened if no numerical estimate for $c_p$ is desired. The left-hand member of (20.2) is a continuous function of the numbers $\beta_j$ in the closed limited point set $|\beta_j|\leqq 1$, hence takes on a minimum value $c_p$; we must prove $c_p>0$. By the hypothesis in the induction, the minimum value zero cannot be taken on when one or several numbers $\beta_j$ vanish, for then by (20.3) the left-hand member of (20.2) equals the corresponding sum with $p$ replaced by some $m<p$ for some function $g_m(z)$: $g_p(z)=z^{p-m}g_m(z)$. The minimum value zero cannot be taken on when all of the numbers $\beta_j$ are different from zero $c_p\geqq|g'(0)|=|\beta_1\beta_2\cdots\beta_{p-1}|>0$. Thus (20.2) is established.

The following series of steps is a consequence of equations (20.5):

$$\begin{aligned} b_1 &= -a_1\alpha, & b_1 &= -a_1\alpha,\\ b_2 - b_1\bar{\alpha} &= a_1 - a_2\alpha, & b_2 - a_1 &= b_1\bar{\alpha} - a_2\alpha,\\ b_3 - b_2\bar{\alpha} &= a_2 - a_3\alpha, & b_3 - a_2 &= b_2\bar{\alpha} - a_3\alpha,\\ &\cdots\cdots\cdots, & &\cdots\cdots\cdots,\\ b_k - b_{k-1}\bar{\alpha} &= a_{k-1} - a_k\alpha, & b_k - a_{k-1} &= b_{k-1}\bar{\alpha} - a_k\alpha,\\ &\cdots\cdots\cdots; & &\cdots\cdots\cdots; \end{aligned}$$

$$\begin{aligned} &| b_2 - a_1 | + | b_3 - a_2 | + \cdots + | b_p - a_{p-1} |\\ &\qquad \leqq [| b_1 | + | b_2 | + \cdots + | b_{p-1} |] \cdot | \alpha |\\ &\qquad\qquad + [| a_2 | + | a_3 | + \cdots + | a_{p-1} |] \cdot | \alpha | + | a_p | \cdot | \alpha |;\\ &[| a_1 | + | a_2 | + \cdots + | a_{p-1} |] - [| b_2 | + | b_3 | + \cdots + | b_p |]\\ &\qquad \leqq [| b_1 | + | b_2 | + \cdots + | b_{p-1} |] \cdot | \alpha |\\ &\qquad\qquad + [| a_2 | + | a_3 | + \cdots + | a_{p-1} |] \cdot | \alpha | + | a_p | \cdot | \alpha |. \end{aligned}$$

Cauchy's inequality for the function $g_{p-1}(z)$ informs us that $|a_p| \leqq 1$, so we may write

$$\begin{aligned} &| b_1 | + | b_2 | + \cdots + | b_p |\\ &\qquad \geqq [| a_1 | + | a_2 | + \cdots + | a_{p-1} |] \frac{1 - |\alpha|}{1 + |\alpha|} - | a_p | \frac{|\alpha|}{1 + |\alpha|}\\ &\qquad \geqq c_{p-1} \frac{1 - |\alpha|}{1 + |\alpha|} - \frac{|\alpha|}{1 + |\alpha|}. \end{aligned} \tag{20.7}$$

*Case* I. $|\alpha| \leqq c_{p-1}/2$. For $p \geqq 2$ we have $c_{p-1} \leqq 1/4$, $|\alpha| \leqq 1/8$; so the last member of (20.7) is not less than

$$c_{p-1}\left[\frac{7}{9} - \frac{1}{2}\right] = \frac{5}{18} c_{p-1} > \frac{c_{p-1}}{2^{p \cdot p!}} = c_p.$$

*Case* II. $|\alpha| > c_{p-1}/2$. Here we replace each term of each of equations (20.6) by the corresponding absolute value. The resulting *inequalities* when added member for member with $k = p - 1$ become (for abbreviation we write $|\alpha| = a$)

$$\begin{aligned} &\left(\frac{1}{a^{p-1}} + \frac{1}{a^{p-2}} - 1\right) | b_1 | + \left(\frac{1}{a^{p-2}} + \frac{1}{a^{p-3}} - 1\right) | b_2 | + \cdots + \frac{1}{a} | b_{p-1} |\\ &\qquad\qquad \geqq | a_1 | + | a_2 | + \cdots + | a_{p-1} | \geqq c_{p-1}. \end{aligned}$$

The coefficient of $|b_1|$ is here not less than the coefficients of $|b_2|$, $|b_3|$, $\cdots$, $|b_{p-1}|$; so we obtain at once from $a > c_{p-1}/2$

$$|b_1| + |b_2| + \cdots + |b_p|$$
$$\geqq \frac{c_{p-1}}{\dfrac{1}{a^{p-1}} + \dfrac{1}{a^{p-2}} - 1} \geqq \frac{a^{p-1}}{2} c_{p-1} \geqq \frac{1}{2}\left(\frac{c_{p-1}}{2}\right)^{p-1} c_{p-1}$$
$$= \left(\frac{c_{p-1}}{2}\right)^p \geqq \frac{1}{2^{(p+1)!}} = c_p.$$

Theorem 3 is completely established.

It is obvious that the choice $c_p = 2^{-(p+1)!}$ can be considerably improved by the present method alone.

It is quite natural to divide the proof of Theorem 3 into two cases depending on the size of $|\alpha|$, comparing $b_j$ with $a_{j-1}$ when $|\alpha|$ is small and comparing $b_j$ with $a_j$ when $|\alpha|$ is large. For it follows from (20.4) that $b_j = a_{j-1}$ when $\alpha = 0$ and that $|b_j| = |a_j|$ when $|\alpha| = 1$.

21. **Numerical upper bound for $D_p$.** Theorem 3, of some interest in itself, is an important step in the proof of Theorem 2. Another preliminary proposition is

THEOREM 4. *Let the function $w = f(z)$ analytic in $|z| < 1$ with $f(0) = 0$ map a smooth region $R$ interior to $|z| = 1$ onto the unit circle $|w| < 1$ covered precisely $p$ times or precisely $m$ times, $m < p$. Then we have*

$$|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)| \geqq \gamma_p > 0, \tag{21.1}$$

*where the number $\gamma_p$ depends on $p$ but not on $R$ or $f(z)$. To be explicit, we shall establish* (21.1) *with* $\gamma_p = 2^{-(p+1)!-p}$.

We shall make use of the analyticity of $f(z)$ only in $R$, not throughout the entire region $|z| < 1$.

Denote by $z = h(Z)$ a function which maps the region $|Z| < 1$ smoothly onto the region $R$ of the $z$-plane, with $h(0) = 0$. Then the function $w = g(Z) = f[h(Z)]$ maps the region $|Z| < 1$ onto the unit circle $|w| < 1$ covered precisely $p$ times or precisely $m < p$ times, with $g(0) = 0$, so $g(Z)$ satisfies the hypothesis of Theorem 3.

Let us introduce the notation

$$g(Z) = a_1 Z + a_2 Z^2 + \cdots,$$
$$f(z) = b_1 z + b_2 z^2 + \cdots,$$
$$h(Z) = d_1 Z + d_2 Z^2 + \cdots.$$

We note that Cauchy's inequality for the function $h(Z)$ yields

$$|d_k| \leqq 1, \qquad k = 1, 2, \cdots. \tag{21.2}$$

The coefficients of $f(z)$ and $g(Z)$ are related by equations that we now need to consider:

$$
\begin{aligned}
g(Z) &= f[h(Z)] \\
&= b_1[d_1Z + d_2Z^2 + d_3Z^3 + \cdots ] \\
&\quad + b_2[d_1Z + d_2Z^2 + d_3Z^3 + \cdots ]^2 \\
&\quad + b_3[d_1Z + d_2Z^2 + d_3Z^3 + \cdots ]^3 \\
&\quad + \cdots\cdots\cdots\cdots\cdots \\
&= a_1Z + a_2Z^2 + a_3Z^3 + \cdots .
\end{aligned} \tag{21.3}
$$

By equating coefficients of corresponding powers of $Z$ we obtain

$$
\begin{aligned}
a_1 &= b_1d_1, \\
a_2 &= b_1d_2 + b_2d_1^2, \\
a_3 &= b_1d_3 + 2b_2d_1d_2 + b_3d_1^3, \\
a_4 &= b_1d_4 + b_2(d_2^2 + 2d_1d_3) + 3b_3d_1^2d_2 + b_4d_1^4, \\
a_5 &= b_1d_5 + b_2(2d_1d_4 + 2d_2d_3) + b_3(3d_1d_2^2 + 3d_1^2d_3) + b_4(4d_1^3d_2) + b_5d_1^5, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots
\end{aligned} \tag{21.4}
$$

The law of the coefficients of the $b_k$ in equations (21.4) is relatively simple, and is readily formulated in terms of the subscripts of the numbers $a_j$ and $b_k$, and involves primarily the partitions of the subscripts of the numbers $a_j$. The precise law would be a needless refinement for our present relatively rough purposes. If we replace each $b_k$ by unity, it is obvious from (21.2) that the function $g(Z)$ in (21.3) is dominated by

$$
\begin{aligned}
&[Z + Z^2 + Z^3 + \cdots ] + [Z + Z^2 + Z^3 + \cdots ]^2 \\
&+ [Z + Z^2 + Z^3 + \cdots ]^3 + \cdots \\
&= \frac{Z}{1 - 2Z} = Z + 2Z^2 + 4Z^3 + 8Z^4 + \cdots .
\end{aligned}
$$

Then the sum of the absolute values of all the coefficients of all the numbers $b_j$ in the first $p$ of equations (21.4) is not greater than $1+2+4+ \cdots +2^{p-1}$, which is less than $2^p$. Insertion in each of equations (21.4) of absolute value signs on the numbers $a_j$, on the numbers $b_j$, and on the coefficients of the numbers $b_j$ yields a corresponding inequality. When the first $p$ of these inequalities are added member for member, there results the inequality

$$
|a_1| + |a_2| + \cdots + |a_p| \leqq 2^p[|b_1| + |b_2| + \cdots + |b_p|],
$$

so (21.1) with $\gamma_p = 2^{-(p+1)!-p}$ is a consequence of Theorem 3.

We are now in a position to prove Theorem 2; the trivial case $D_p(0) = 0$ needs no further discussion and is henceforth excluded. Under the hypothesis

of Theorem 2 the function

$$w_1(z) = \frac{f(z)}{D_p(0)} \tag{21.5}$$

is analytic for $|z| < 1$ and maps a smooth region $R$ interior to $|z| = 1$ onto the region $|w_1| < 1$ covered precisely $p$ times or precisely $m < p$ times, with $w_1(0) = 0$. Theorem 4 applied to the function (21.5) yields at once inequality (20.1). Theorem 2 is established, and we may state the

COROLLARY. *In Theorem 2 we may take* $\gamma_p = 2^{-(p+1)!-p}$.

The number $2^{-(p+1)!-p}$ can obviously be greatly improved, even without change of method.

Theorem 2 is stated in the form convenient for applications, but we have used in the proof the analyticity of $f(z)$ not in the entire region $|z| < 1$, only in a neighborhood of the origin. However, if $f(z)$ is analytic in a region containing points for which $|z| \geqq 1$, the number $D_p(0)$ is to be defined as referring to the Riemann configuration which is the image of $|z| < 1$ under the transformation $w = f(z)$. Theorem 2 is false if the points $|z| \geqq 1$ are not excluded, as is shown by the example $p = 1$, $f(z) \equiv z$.

It is clear now that from Theorem 1 (with Corollary 1) and Theorem 2 of the present chapter, Theorem 19 of Chapter II may be obtained in the explicit form of inequalities. Indeed, we have

THEOREM 5. *Let $f(z)$ be regular in $|z| < 1$ and bounded there:*

$$|f(z)| < M.$$

*Let $\{z_n\}$ ($|z_n| < 1$) be a sequence of points in $|z| < 1$ and let $w_n = f(z_n)$. Then, there exist two constants $\lambda_p$ and $\Lambda_p$ of which $\lambda_p$ depends on $p$ alone, while $\Lambda_p$ depends on $p$ and $M$ so that*

$$\begin{aligned} \lambda_p \cdot D_p(w_n) &\leqq \sum_{k=1}^{p} \left| \sum_{\nu=0}^{k-1} (-1)^{\nu} C_{k-1,\nu} \bar{z}_n^{\nu} \frac{(1-|z_n|^2)^{k-\nu} f^{(k-\nu)}(z_n)}{(k-\nu)!} \right| \\ &\leqq \Lambda_p [D_p(w_n)]^{2^{-p}}, \end{aligned} \tag{21.6}$$

*where $D_p(w_n)$ is the radius of $p$-valence at the point $w_n$ of the Riemann surface on which $w = f(z)$ maps the circle $|z| < 1$.*

The writers are not informed as to whether the exponent $2^{-p}$ in (21.6) is the best possible one. Here, and in improving the constants $\lambda_p$ and $\Lambda_p$ already obtained, lie a number of interesting open problems.

As a consequence of the second half of inequality (21.6) and the example of §9, Theorem 8 we may state

THEOREM 6. *Let the function $Q(r)$ be defined and positive for $0 < r < 1$, with* $\lim_{r\to 1} Q(r) = 0$. *Let the positive integer $m$ be given. Then there exist a function*

*$w=f(z)$ analytic and univalent in $|z|<1$, continuous in $|z|\leqq 1$, and a sequence of points $z_1, z_2, \cdots$ with $|z_n|<1$, $|z_n|\to 1$, such that we have*

$$\lim_{n\to\infty}\frac{D_m(w_n)}{Q(|z_n|)}=\infty,$$

*where $w_n=f(z_n)$.*

## CHAPTER IV. FUNCTIONS WHICH OMIT TWO VALUES

22. **Inequalities for $D_p(w_n)$ when $|f(z_n)|$ is bounded.** Practically all the results of the Chapters II and III may be extended to the class of functions $f(z)$ regular in the circle $|z|<1$ which in that circle differ from 0 and 1[57]. To be more specific, suppose that $f(z)$ is regular in the circle $|z|<1$ and that $f(z)\neq 0$, 1 in $|z|<1$. Let $\{z_n\}$ ($|z_n|<1$) be an arbitrary sequence of points in the circle so that $|f(z_n)|$ remains bounded for all $n$. Under these assumptions what is a necessary and sufficient condition that $(1-|z_n|)^k f^{(k)}(z_n)\to 0$ ($k=1, 2, \cdots, p$)? If we examine the proof of Theorem 19, Chapter II, we notice that absolutely no modification is necessary in order to extend this theorem to the case under consideration since we are again dealing with a normal family $\{\phi_n(\zeta)\}$ which, due to the condition that $|f(z_n)|$ is bounded, does not contain the infinite constant. The proof of Theorem 19, therefore, may be repeated verbatim to yield

THEOREM 1. *Let $f(z)$ be regular in $|z|<1$ and $f(z)\neq 0$, 1 there. Let $\{z_n\}$ ($|z_n|<1$) be a sequence of points in $|z|<1$ such that $|f(z_n)|<M$ for all $n$. Then, a necessary and sufficient condition that*

$$\lim_{n\to\infty}|f^{(k)}(z_n)|(1-|z_n|)^k=0,\qquad k=1\ 2\ \cdots, p,$$

*for a fixed positive integer $p$ is that*

$$\lim_{n\to\infty}D_p(w_n)=0,$$

*where $w_n=f(z_n)$.*

Again as in the case of Theorem 19 it is desirable to give explicitly the relation between $|f^{(p)}(z_n)|(1-|z_n|^2)^p$ and $D_p(w_n)$. In view of §21, Theorem 5 and Schottky's theorem this relation is easily obtained. We use Schottky's theorem in the following form[58]: *If $f(z)$ is regular in $|z|<1$ and omits there the values zero and one, if $f(z)=a_0+a_1z+\cdots$, then there exists a positive constant $\Delta$, independent of $a_0$, $\theta$, $a_1, \cdots$ so that*

$$(22.1)\qquad |f(z)|<[|a_0|+2]^{\Delta/(1-\theta)}$$

*in the circle $|z|<\theta<1$.*

---

(57) The case $f(z)\neq a$, $b$ may always be reduced to the above case by considering $\phi(z)=(f(z)-a)/(b-a)$.

(58) Cf. L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, 2d. edition, 1931, p. 224.

Let us assume now that the hypotheses of Theorem 1 are satisfied, and form the functions

$$\phi_n(\zeta) = f\left(\frac{\zeta + z_n}{1 + \bar{z}_n\zeta}\right). \tag{22.2}$$

These functions are all regular in $|\zeta| < 1$ and omit there the two values 0 and 1. Furthermore, $\phi_n(0) = f(z_n) = w_n$ are bounded in absolute value by the constant $M$:

$$|\phi_n(0)| < M, \qquad n = 1, 2, \cdots.$$

Applying Schottky's theorem in the form (22.1) to the functions $\phi_n(\zeta)$, we find that

$$|\phi_n(\zeta)| < [M + 2]^{\Delta/(1-\theta)} = M_\theta$$

in the circle $|\zeta| < \theta < 1$. If we set now

$$g_n(\zeta) = \phi_n(\theta\zeta), \tag{22.3}$$

we obtain a regular function $g_n(\zeta)$ in the circle $|\zeta| < 1$ which satisfies the inequality $|g_n(\zeta)| < M_\theta$ in the whole circle $|\zeta| < 1$. Finally we set

$$h_n(\zeta) = g_n(\zeta) - g_n(0), \tag{22.4}$$

so that $h_n(\zeta)$ is regular in $|\zeta| < 1$, $h_n(0) = 0$, and

$$|h_n(\zeta)| < 2M_\theta.$$

Now, according to §21, Theorem 5, we have

$$\lambda_p \cdot D_p(0) \leqq |h_n'(0)| + \frac{1}{2!}|h_n''(0)| + \cdots + \frac{1}{p!}|h_n^{(p)}(0)| \leqq \Lambda_p \cdot [D_p(0)]^{2-p},$$

where $D_p(0)$ is the radius of $p$-valence (§14) at the point $w = 0$ of the Riemann configuration $R_n$ on which $w = h_n(z)$ maps the circle $|z| < 1$. From (22.4) we obtain

$$\lambda_p \cdot D_p(0) \leqq |g_n'(0)| + \frac{1}{2!}|g_n''(0)| + \cdots + \frac{1}{p!}|g_n^{(p)}(0)| \leqq \Lambda_p \cdot [D_p(0)]^{2-p}.$$

But $g_n(\zeta)$ maps $|\zeta| < 1$ on a Riemann configuration $R_n'$ obtained from $R_n$ by translating it along the vector $g_n(0)$. Therefore, $D_p(0)$ is equal to the radius of $p$-valence of $R_n'$ at the point $w = g_n(0)$. This radius we shall denote by $D_p[g_n(0)]$. We thus obtain

$$\lambda_p \cdot D_p[g_n(0)] \leqq |g_n'(0)| + \frac{1}{2!}|g_n''(0)| + \cdots + \frac{1}{p!}|g_n^{(p)}(0)|$$
$$\leqq \Lambda_p [D_p[g_n(0)]]^{2-p}.$$

By virtue of (22.3) this becomes

$$\begin{aligned}(22.5)\qquad \lambda_p\cdot D_p[g_n(0)] &\leqq \theta\,|\,\phi_n'(0)\,| + \frac{\theta^2}{2!}\,|\,\phi_n''(0)\,| + \cdots + \frac{\theta^p}{p!}\,|\,\phi_n^{(p)}(0)\,| \\ &\leqq \Lambda_p\cdot [D_p[g_n(0)]]^{2-p}.\end{aligned}$$

Now, the Riemann surface $R_n'$ can simply be considered as the surface on which the function $w=\phi_n(\zeta)$ maps the circle $|\zeta|<\theta$. It is, therefore, merely a part of the surface $R$ on which $\phi_n(\zeta)$, and by (22.2) $w=f(z)$, maps the circle $|z|<1$. If we denote by $D_p(w_n)$ the radius of $p$-valence of $R$ at the point $w=w_n$, we clearly must have

$$D_p[g_n(0)] \leqq D_p(w_n).$$

We may, therefore, infer the inequality

$$\theta\,|\,\phi_n'(0)\,| + \frac{\theta^2}{2!}\,|\,\phi_n''(0)\,| + \cdots + \frac{\theta^p}{p!}\,|\,\phi_n^{(p)}(0)\,| \leqq \Lambda_p\cdot [D_p(w_n)]^{2-p}.$$

Now, since $0<\theta<1$, we find

$$|\,\phi_n'(0)\,| + \frac{1}{2!}\,|\,\phi_n''(0)\,| + \cdots + \frac{1}{p!}\,|\,\phi_n^{(p)}(0)\,| \leqq \frac{\Lambda_p}{\theta^p}\,[D_p(w_n)]^{2-p}.$$

According to (22.2) and (2.3) we obtain

$$(22.6)\qquad \sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^\nu C_{k-1,\nu}\bar{z}_n^\nu\,\frac{(1-|\,z_n\,|^2)^{k-\nu}f^{(k-\nu)}(z_n)}{(k-\nu)!}\right| \leqq \frac{\Lambda_p}{\theta^p}\,[D_p(w_n)]^{2-p}.$$

This gives us the desired inequality from above. The corresponding inequality from below, is contained in §20, Theorem 2:

$$(22.7)\qquad \lambda_p\cdot D_p(w_n) \leqq \sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^\nu C_{k-1,\nu}\bar{z}_n^\nu\,\frac{(1-|\,z_n\,|^2)^{k-\nu}f^{(k-\nu)}(z_n)}{(k-\nu)!}\right|.$$

We may, therefore, state the following

THEOREM 2. *Let $f(z)$ be regular in $|z|<1$ and $f(z)\neq 0$, 1 there. Let $\{z_n\}$ ($|z_n|<1$) be a sequence of points in $|z|<1$ such that $|f(z_n)|<M$ for all $n$. Then, for any $0<\theta<1$ there exist two constants $\lambda_p$ and $\Lambda_p$ of which $\lambda_p$ depends on $p$ alone, while $\Lambda_p$ depends on $p$, $M$, and $\theta$, so that*

$$\begin{aligned}(22.8)\qquad \lambda_p\cdot D_p(w_n) &\leqq \sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^\nu C_{k-1,\nu}\bar{z}_n^\nu\,\frac{(1-|\,z_n\,|^2)^{k-\nu}f^{(k-\nu)}(z_n)}{(k-\nu)!}\right| \\ &\leqq \frac{\Lambda_p}{\theta^p}\,[D_p(w_n)]^{2-p},\end{aligned}$$

*where $D_p(w_n)$ is the radius of p-valence at the point $w_n=f(z_n)$ of the Riemann surface on which $w=f(z)$ maps the circle $|z|<1$.*

Since from the form of $\Lambda_p$ it is evident that it tends to infinity as $\theta$ tends to 1, the best value for the right side of (22.8) is obtained for that value of $\theta$ for which $\Lambda_p/\theta^p$ attains its minimum. That value may be readily computed from the expression for $\Lambda_p$. It is evident also that Theorem 2 implies Theorem 1.

We remark that under the conditions of Theorem 2 we have $D_p(w)\leqq|w|$, so that (22.8) gives an inequality on the approach to zero of $(1-|z|^2)^kf^{(k)}(z)$ as $w$ tends to zero, for every $k$.

A further consequence of Theorem 2 is that under the hypothesis of that theorem, an additional inequality of the form $|f(z)|\leqq M$ implies inequalities $|f^{(k)}(z)|(1-|z|^2)^k\leqq M_k$, where $M_k$ depends only on $k$ and $M$. Indeed, we have $D_p(w)\leqq M$; our conclusion[59] follows from (22.8).

23. **Counterexamples.** In Theorems 1 and 2 an important part of the hypothesis was the fact that $|f(z_n)|<M$ for all $n$. Since any sequence $\{z_n\}$ can be decomposed into sequences on which $|f(z_n)|$ is bounded and those on which $|f(z_n)|$ tends to infinity, it is natural to inquire how far Theorem 2 can be extended to sequences $\{z_n\}$ for which $|f(z_n)|\to\infty$.

That the conclusion of Theorem 2 as a proposition is false for such sequences is a theorem which we shall establish:

THEOREM 3. *There exists a function $f(z)$ with two omitted values and regular in $|z|<1$ and there exists a sequence of points $\{z_n\}$ ($|z_n|<1$, $|z_n|\to 1$) such that, setting $w_n=f(z_n)$, we have $D_1(w_n)\to 0$, $w_n\to\infty$, and yet $\lim_{n\to\infty}|f'(z_n)|(1-|z_n|^2)=8\pi$.*

In the half-plane $\Re W>0$, where $W=u+iv$,

$$\Re(W+e^{-W+1})=u+e^{-u+1}\cos v\geqq -e^{-u+1}\geqq -e.$$

Consequently, in $\Re W>0$ the function $W+e^{-W+1}+3$ omits all values in some neighborhood of the origin, as does the function

$$w=f(z)=(W+e^{-W+1}+3)^2=F(W), \tag{23.1}$$

where we set $W=(1+z)/(1-z)$, so that $z$ is a point of the unit circle $|z|<1$. We choose $W_n=1+2n\pi i+1/n$, whence $e^{-W_n+1}=e^{-1/n}$ and find

$$\frac{df(z)}{dW}=2(W+e^{-W+1}+3)(1-e^{-W+1}). \tag{23.2}$$

Thus, $f'(z)$ vanishes in the points where $1-e^{-W+1}=0$, namely $W=1+2n\pi i$, $n=0,\pm1,\pm2,\cdots$. If we define $z_n$ by the relation $W_n=(1+z_n)/(1-z_n)$, we find from (23.2)

(59) More precise inequalities of this type were developed by O. Szász, loc. cit.

$$\frac{df(z_n)}{dW} = 2\left(4 + 2n\pi i + \frac{1}{n} + e^{-1/n}\right)(1 - e^{-1/n}),$$

so that $df(z_n)/dW \to 4\pi i$. We next compute $|1-z|^2 = 4/|W+1|^2$ and $|dW/dz| = |W+1|^2/2$. Hence,

$$\left|\frac{dW}{dz}\right|_{z=z_n}(1 - |z_n|^2) = \frac{1}{2}\left[|W_n + 1|^2 - |W_n - 1|^2\right] = 2 + \frac{2}{n} \to 2.$$

Thus, we obtain finally

$$|f'(z_n)|(1 - |z_n|^2) \to 8\pi.$$

It now remains to be shown that $D_1(w_n) \to 0$. This may be shown as follows. In the $W$-plane consider the two points $W = 1 + 2n\pi i$ and $W_n = 1 + 2n\pi i + 1/n$. Join these two points by a rectilinear segment, necessarily horizontal. This segment is mapped by the function $w = F(W)$ on a certain arc lying on the corresponding Riemann configuration and joining the points $w = (5 + 2n\pi i)^2$ and $w_n = (4 + e^{-1/n} + 1/n + 2n\pi i)^2$, of which the first is a branch point of the Riemann configuration in question. It is clear, therefore, that this arc emanates from the center of the circle $|w - w_n| \leqq D_1(w_n)$ and terminates in a point lying exterior to or on the boundary of that circle([60]). Hence, the length of this arc cannot be less than $D_1(w_n)$. But the length can be estimated directly. Indeed, it is equal to

$$\int_1^{1+1/n} |F'(2n\pi i + u)|\, du.$$

From (23.2) we find

$$D_1(w_n) \leqq 2\int_1^{1+1/n} |2n\pi i + u + e^{-u+1} + 3|(1 - e^{-u+1})du.$$

Now, in the interval $1 \leqq u \leqq 1 + 1/n$, we have $1 - e^{-u+1} \leqq 1 - e^{-1/n}$ and $e^{-u+1} \leqq 1$, so that

$$D_1(w_n) \leqq 2(1 - e^{-1/n})(2n\pi + 5 + 1/n)\cdot 1/n. \tag{23.3}$$

Hence, as $n \to \infty$, we have $D_1(w_n) \to 0$, which completes the proof of the theorem.

In connection with the present example one may make two remarks.

*Remark* 1. If one replaces the function $f(z)$ in (23.1) by the function

$$f(z) = (W + e^{-W+1} + 3)^4, \qquad W = (1 + z)/(1 - z), \tag{23.4}$$

([60]) Study of the variation of arg $(dw)$ on the arc shows that the arc lies wholly in the circle in question, and hence that $D_1(w_n) = |F(W) - F(W_n)|$, where $W = 1 + 2n\pi i$. A similar fact holds under Remarks 1 and 2.

with $W_n+1=2n\pi i+1/n^2$, clearly the relation $D_1(w_n)\to 0$ still holds, while $|f'(z_n)|(1-|z_n|^2)\to\infty$. Thus, $D_1(w_n)\to 0$ *does not even imply the boundedness* of $|f'(z_n)|\ (1-|z_n|^2)$.

*Remark* 2. Let $\alpha$ be any real number in the interval $0<\alpha<1$. Choose an integer $k$ so that $k>\alpha/(1-\alpha)$. Then, the choice

$$f(z) = (W + e^{-W+1} + 3)^{k+1}, \qquad W = (1+z)/(1-z),$$

with $W_n=1+2n\pi i+1/n^k$ yields $D_1(w_n)\to 0$. Indeed, a computation analogous to the one in the preceding example shows that $D_1(w_n)=O(1/n^k)$. On the other hand, $|w_n|=O(n^{k+1})$. Hence $|w_n|^\alpha\cdot D_1(w_n)=O(n^{\alpha k+\alpha-k})$ and this expression tends to zero. Furthermore, it is easily seen that $|f'(z_n)|\ (1-|z_n|^2)>c>0$, where $c$ is a certain positive constant. Thus, *for the class of functions with a region of omitted values no relation* $|w_n|^\alpha\cdot D_1(w_n)\to 0$ *with* $0<\alpha<1$ *can imply* $|f'(z_n)|\ (1-|z_n|^2)\to 0$.

In the example of Theorem 3 and the examples in the two remarks it will be noticed that $|w_n|\cdot D_1(w_n)$ does not tend to zero. The case that $|w_n|\cdot D_1(w_n)$ tends to zero will not be treated in its full generality in this paper. A special case is considered in §25. The case $|w_n|^\alpha\cdot D_1(w_n)\to 0$ for $\alpha>1$ will be considered in the next section.

24. **Case:** $\lim_{n\to\infty} |w_n|^{(1+\epsilon)(2^p-1)}D_p(w_n)=0$. The following extension of Theorem 1 for $p=1$ to the case $|w_n|\to\infty$ will now be proved:

THEOREM 4. *Let $f(z)$ be analytic in $|z|<1$ and omit two values there. Let $\{z_n\}$ ($|z_n|<1$) be a sequence of points in $|z|<1$ such that, setting $w_n=f(z_n)$, we have* $\lim_{n\to\infty}|w_n|=\infty$. *Then, the condition*

$$\lim_{n\to\infty} |w_n|^{1+\epsilon}D_1(w_n) = 0 \tag{24.1}$$

*for any positive* $\epsilon$ *implies*

$$\lim_{n\to\infty} |f'(z_n)|\,(1-|z_n|^2) = 0. \tag{24.2}$$

It is clear that the sequence of functions

$$\phi_n(\zeta) = f\left(\frac{\zeta+z_n}{1+\bar{z}_n\zeta}\right)$$

regular in $|\zeta|<1$ is normal. Since by hypothesis $\lim_{n\to\infty}|w_n|=\infty$, we have

$$\lim_{n\to\infty} |\phi_n(0)| = \infty,$$

so that

$$\lim_{n\to\infty} |\phi_n(\zeta)| = \infty \tag{24.3}$$

uniformly in every closed subregion of $|\zeta|<1$.

Choose a positive number

$$\rho < \frac{\epsilon}{2+\epsilon} < 1,$$

whence

$$\frac{1+\rho}{1-\rho} < 1+\epsilon.$$

It follows from (24.3) that for $n$ sufficiently large the function $1/\phi_n(\zeta)$ is regular in the circle $|\zeta| \leqq \rho_1$, where $\rho_1$ is any number such that $\rho < \rho_1 < 1$. Furthermore, $n$ may be chosen so large that $1/|\phi_n(\zeta)| < 1$ in $|\zeta| \leqq \rho_1$, which implies that $\log |\phi_n(\zeta)|$ is harmonic and positive in $|\zeta| \leqq \rho_1$. Then, using Poisson's integral for the region $|\zeta| < \rho_1$, one sees immediately that

$$\log |\phi_n(\zeta)| \leqq \frac{\rho_1 + \rho}{\rho_1 - \rho} \log |\phi_n(0)| \tag{24.4}$$

in the circle $|\zeta| \leqq \rho < \rho_1$. Now, by taking $\rho_1$ so near to unity that $\rho_1 + \rho/\rho_1 - \rho < 1+\epsilon$ and then by choosing $n$ sufficiently large, the inequality (24.4) implies $|\phi_n(\zeta)| < |\phi_n(0)|^{1+\epsilon}$ in the circle $|\zeta| \leqq \rho$[61].

Now, according to Theorem 3 of Chapter II,

$$D_1(w_n) \geqq \frac{|\phi_n'(0)|^2 r^2}{8M_n},$$

where $M_n \geqq \max_{|\zeta| \leqq r} |\phi_n(\zeta)|$. If we set $r = \rho$, $M_n = |\phi_n(0)|^{1+\epsilon}$, $\phi_n(0) = w_n$, we obtain for $n$ sufficiently large

$$D_1(w_n) \geqq \frac{|f'(z_n)|^2 (1 - |z_n|^2)^2 \rho^2}{8|w_n|^{1+\epsilon}}, \tag{24.5}$$

from which the theorem follows at once.

The treatment of the case for general $p$ is quite analogous:

THEOREM 5. *Let $f(z)$ be analytic in $|z| < 1$ and omit two values there. Let $\{z_n\}$ ($|z_n| < 1$) be a sequence of points in $|z| < 1$ such that, setting $w_n = f(z_n)$, $\lim_{n\to\infty} |w_n| = \infty$. Then, the condition*

$$\lim_{n\to\infty} |w_n|^{(1+\epsilon)(2^p-1)} D_p(w_n) = 0 \tag{24.6}$$

*for any positive $\epsilon$ implies*

$$\lim_{n\to\infty} |f^{(j)}(z_n)| (1 - |z_n|^2)^j = 0, \qquad j = 1, 2, \cdots, p. \tag{24.7}$$

---

(61) The reasoning employed in the proof of this inequality is well known. Cf. A. Ostrowski, Abhandlungen des Mathematischen Seminars der Hamburgischen Universität, vol. 1 (1922), pp. 327–350; S. Mandelbrojt, Comptes Rendus de l'Académie des Sciences, Paris, vol. 185 (1927), pp. 1098–1100; H. Cartan, Annales de l'École Normale Supérieure, (3), vol. 45 (1928), pp. 255–346; J. L. Walsh, Tôhoku Mathematical Journal, vol. 38 (1933), pp. 375–389.

The proof of Theorem 4 is repeated verbatim, and we find, as before, that for any positive $\epsilon$ there exists a positive number $\rho<1$ such that for all sufficiently large values of $n$

$$|\phi_n(\zeta)| < |w_n|^{1+\epsilon}$$

in the circle $|\zeta| \leqq \rho$.

Now, according to inequality (19.5′), we obtain

$$\sum_{j=1}^{p} \frac{\rho^j}{j!} | \phi_n^{(j)}(0) | \leqq 24 | w_n |^{1+\epsilon} \sum_{j=1}^{p} \left( \frac{D_j(w_n)}{| w_n |^{1+\epsilon}} \right)^{1/2^j}$$

$$= 24 \sum_{j=1}^{p} ( | w_n |^{(1+\epsilon)(2^j-1)} D_j(w_n))^{1/2^j},$$

whence, applying (2.3), we find for $n$ sufficiently large

$$\sum_{j=1}^{p} \rho^j \left| \sum_{\nu=0}^{j-1} (-1)^\nu C_{j-1,\nu} \bar{z}_n^\nu \frac{(1 - | z_n |^2)^{j-\nu} f^{(j-\nu)}(z_n)}{(j-\nu)!} \right|$$

$$\leqq 24 \sum_{j=1}^{p} ( | w_n |^{(1+\epsilon)(2^j-1)} D_j(w_n))^{1/2^j}.$$

Since (24.6) implies the relation

$$\lim_{n\to\infty} | w_n |^{(1+\epsilon)(2^j-1)} D_j(w_n) = 0, \qquad j = 1, 2, \cdots, p,$$

we obtain (24.7).

25. **Mandelbrojt's theorem.** The following theorem is due to S. Mandelbrojt[62]:

THEOREM A. *Let $f_n(z)$ be a sequence of functions analytic in a region $R$ and tending uniformly in $R$ to infinity. If there exists a positive constant $M$ such that for all $n$ and for all $z$ in $R$*

(25.1) $$| \arg f_n(z) | < M$$

*with some determination of the argument, then to every closed region $R_1$ wholly interior to $R$ there corresponds a finite positive number $\alpha$ $(1<\alpha<+\infty)$ and a positive integer $n_0$ such that for every pair of points $z_0$ and $z_1$ in $R_1$ and for every $n>n_0$, the inequality*

(25.2) $$\frac{1}{\alpha} < \left| \frac{f_n(z_1)}{f_n(z_0)} \right| < \alpha$$

*holds.*

---

(62) S. Mandelbrojt, loc. cit.

We indicate a proof of Theorem A[63]. Let us first prove the assertion of the theorem in the special case that $R_1$ is the circle $C: |z-a| \leqq \rho$ lying wholly in $R$. It is clear by hypothesis that for sufficiently large values of $n$ the functions $f_n(z) \neq 0$ in $C$, and henceforth we shall consider only such values of $n$. Hence, the functions $-i \log f_n(z)$ will be regular in $C$, single-valued in $C$ after a particular determination of the logarithm is selected. We choose that determination for which $\Re[-i \log f_n(z)] = \arg f_n(z)$, where the argument is the one asserted in (25.1). Now, take a circle $C'$: $|z-a| \leqq \rho'$ for which $\rho' > \rho$ and which also lies wholly in $R$; choose $n$ so large that $f_n(z) \neq 0$ in $C'$.

In $C'$ we have the representation

$$(25.3)\qquad \begin{aligned} &- \log | f_n(a + re^{i\theta}) | + \log | f_n(a) | \\ &= \frac{1}{2\pi} \int_0^{2\pi} \arg f_n(a + \rho' e^{i\phi}) \frac{2\rho' r \sin (\theta - \phi)}{\rho'^2 + r^2 - 2\rho' r \cos (\theta - \phi)} d\phi. \end{aligned}$$

Let $z_0 = a + r_0 e^{i\theta_0}$ and $z_1 = a + r_1 e^{i\theta_1}$ be any two points of $C$. We may, then, write (25.3) for the points $z_0$ and $z_1$ and subtract the second equation from the first. Thus, we obtain the equation

$$(25.4)\qquad \begin{aligned} \log \left| \frac{f_n(z_1)}{f_n(z_0)} \right| = \frac{1}{\pi} \int_0^{2\pi} \arg f_n(a + \rho' e^{i\phi}) \Bigg[ & \frac{\rho' r_0 \sin (\theta_0 - \phi)}{\rho'^2 + r_0^2 - 2\rho' r_0 \cos (\theta_0 - \phi)} \\ & - \frac{\rho' r_1 \sin (\theta_1 - \phi)}{\rho'^2 + r_1^2 - 2\rho' r_1 \cos (\theta_1 - \phi)} \Bigg] d\phi. \end{aligned}$$

Taking absolute values in (25.4) and observing (25.1), we find

$$\begin{aligned} &\left| \log \left| \frac{f_n(z_1)}{f_n(z_0)} \right| \right| \\ &\leqq \frac{M}{\pi} \int_0^{2\pi} \left| \frac{\rho' r_0 \sin (\theta_0 - \phi)}{\rho'^2 + r_0^2 - 2\rho' r_0 \cos (\theta_0 - \phi)} - \frac{\rho' r_1 \sin (\theta_1 - \phi)}{\rho'^2 + r_1^2 - 2\rho' r_1 \cos (\theta_1 - \phi)} \right| d\phi. \end{aligned}$$

But since $z_0$ and $z_1$ lie in the circle $C$, an easy calculation shows that

$$(25.5)\qquad \left| \log \left| \frac{f_n(z_1)}{f_n(z_0)} \right| \right| \leqq \frac{4\rho\rho' M}{(\rho' - \rho)^2}.$$

The right-hand side in (25.5) is independent of the pair of points $z_0$ and $z_1$. From (25.5) follows at once the assertion of Theorem A in the case that $R_1$ is a circle.

We now pass to the general case. Let $R_1'$ be any closed region contained in $R$ and itself containing $R_1$ in its interior. Consider the class of all open

[63] The proof of this theorem given by Mandelbrojt, loc. cit., is not clear to the writers. The proof given in the text was suggested to the authors by Professor S. E. Warschawski.

circles with centers in $R_1'$ contained together with their boundaries in $R$. In accordance with the Heine-Borel theorem one may select out of this class a finite number of circles which cover $R_1'$. Denote this number by $N$. By the first part of the proof with each one of these circles there is associated a number $\alpha_\nu$ $(1<\alpha_\nu<\infty)$ and a positive integer $\lambda_\nu$, such that for any pair of points $z_0$ and $z_1$ in that circle and for $n>\lambda_\nu$

$$\frac{1}{\alpha_\nu} < \left| \frac{f_n(z_1)}{f_n(z_0)} \right| < \alpha_\nu.$$

Let $\beta$ and $n_0$ be the largest of the numbers $\alpha_\nu$ and $\lambda_\nu$, respectively. Then, for $n>n_0$ and for any pair of points in any one of those circles we have

$$\frac{1}{\beta} < \left| \frac{f_n(z_1)}{f_n(z_0)} \right| < \beta. \tag{25.6}$$

Now consider any two points $z_0$ and $z_1$ in $R_1$. Connect $z_0$ and $z_1$ by a simple polygonal line $P$ lying wholly in $R_1'$ and so chosen as not to be tangent to any circle of the above class. Denote by $C_1$ any circle of the above class which contains the point $z_0$. As one travels along $P$ from $z_0$ to $z_1$, there will be a last point of intersection $\zeta_1$ of $P$ with the circumference of $C_1$. Denote by $C_2$ any circle of the above class which contains $\zeta_1$. Between $z_0$ and $\zeta_1$ on $P$ choose any point $\xi_1$ common to both $C_1$ and $C_2$. Now, starting with the point $\xi_1$ which belongs to $C_2$, repeat the argument. We obtain in this manner a point $\xi_2$ of $P$ which is common to two circles $C_2$ and $C_3$ of the above family. Proceeding in this manner, after a finite number of steps we come to a first circle $C_k$ which contains the point $z_1$. It is clear from (25.6) that for $n>n_0$

$$\frac{1}{\beta^N} \leqq \frac{1}{\beta^k} < \left| \frac{f_n(z_1)}{f_n(z_0)} \right| = \left| \frac{f_n(\xi_1)}{f_n(z_0)} \right| \cdot \left| \frac{f_n(\xi_2)}{f_n(\xi_1)} \right| \cdots \left| \frac{f_n(z_1)}{f_n(\xi_{k-1})} \right| < \beta^k \leqq \beta^N.$$

Setting $\beta^N=\alpha$, we obtain the constant asserted in Theorem A.

As Mandelbrojt himself points out, these results may be readily extended to the case of a sequence of functions $f_n(z)$ regular in $R$ which converges uniformly in $R$ to an analytic function $f(z)$ in such a manner that the differences $f_n(z)-f(z)$ do not vanish in $R$.

Theorem A may be used to obtain a result related to Theorem 4.

THEOREM 6. *Let $f(z)$ be analytic in $|z|<1$ and omit two values there, including the value $w=a$. Let $\{z_n\}$ $(|z_n|<1)$ be a sequence of points in $|z|<1$ such that, setting $w_n=f(z_n)$, we have $\lim_{n\to\infty}|w_n|=\infty$. If $\arg[f(z)-a]$ is uniformly bounded in $|z|<1$*(64)*, then the condition*

$$\lim_{n\to\infty} |w_n| \cdot D_1(w_n) = 0$$

(64) Geometrically, this condition means that the Riemann surface on which $w=f(z)$ maps the unit circle $|z|<1$ does not wind infinitely many times about the point $w=a$.

*implies the relation*

$$\lim_{n\to\infty} |f'(z_n)|(1-|z_n|) = 0.$$

The boundedness of arg $[f(z)-a]$ implies the boundedness in $|\zeta|<1$ of arg $[\phi_n(\zeta)-a]$, where

$$\phi_n(\zeta) = f\left(\frac{\zeta+z_n}{1+\bar{z}_n\zeta}\right).$$

Just as in the proof of Theorem 4 we infer that

$$\lim_{n\to\infty} |\phi_n(\zeta)| = \infty$$

uniformly in every closed subregion of $|\zeta|<1$. Hence,

$$\lim_{n\to\infty} |\phi_n(\zeta)-a| = \infty$$

uniformly in every closed subregion of $|\zeta|<1$. We may therefore apply Theorem A of Mandelbrojt to the sequence of functions $\phi_n(\zeta)-a$ in the circle $|\zeta|<\rho$, where $\rho$ is any fixed positive number less than unity. It follows that corresponding to any circle $|\zeta|\leqq\rho_1<\rho$ one may assign a finite positive number $\alpha$ and a positive integer $n_0$ such that for any pair of points $\zeta$, $\zeta_0$ in $|\zeta|\leqq\rho_1$

$$\frac{1}{\alpha} < \left|\frac{\phi_n(\zeta)-a}{\phi_n(\zeta_0)-a}\right| < \alpha$$

for every $n>n_0$. In particular, choosing $\zeta_0=0$, we obtain the inequality

$$|\phi_n(\zeta)-a| < \alpha|\phi_n(0)-a|$$

and

$$(25.7)\qquad |\phi_n(\zeta)| < \alpha|\phi_n(0)| + (\alpha+1)|a| = \alpha|w_n| + (\alpha+1)|a|$$

in $|\zeta|\leqq\rho_1$ provided $n>n_0$.

Thus, one may apply Theorem 3 of Chapter II where $M=M_n=\alpha|w_n| +(\alpha+1)|a|$. The theorem follows at once. Conditions more delicate than those given in Theorems 4 and 5 may be obtained by different methods. Thus, it may be shown that the condition (24.1) may be replaced by the less stringent condition

$$\lim_{n\to\infty} |w_n|(\log|w_n|)^{1+\epsilon}D_1(w_n) = 0.$$

This result, and other analogous ones, will be developed in a later joint paper of A. S. Galbraith, W. Seidel, and J. L. Walsh.

26. **Counterexample for unrestricted functions.** In obtaining relations between $|f'(z)|(1-|z|)$ and $D_1(w)$ we have always restricted the class of func-

tions $f(z)$. We have thus far considered univalent functions, bounded functions, and functions omitting two values. That these or similar restrictions are essential is shown by the following example.

THEOREM 7. *There exists a function $f(z)$ analytic in the unit circle $|z|<1$ and a sequence of points $\{z_n\}$ ($|z_n|<1$, $|z_n|\to 1$) such that, setting $w_n=f(z_n)$, we have $D_1(w_n)\to 0$, $|w_n|$ bounded, and*

$$\lim_{n\to\infty} |f'(z_n)| (1-|z_n|) = 4\pi.$$

Consider the function

$$w = f(z) = \sin^2 W,$$

where $W=(1+z)/(1-z)$. It follows that

$$f'(z)(1-z) = \frac{2 \sin 2W}{1-z}\cdot$$

Let us set

$$z_n = \frac{1/n + 2n\pi - 1}{1/n + 2n\pi + 1}, \qquad \zeta_n = \frac{2n\pi - 1}{2n\pi + 1}, \qquad W_n = \frac{1}{n} + 2n\pi.$$

We find

$$f'(z_n)(1-z_n) = \left(1 + \frac{1}{n} + 2n\pi\right) \sin \frac{2}{n},$$

$$\lim_{n\to\infty} f'(z_n)(1-z_n) = 4\pi.$$

On the other hand, setting $w_n=f(z_n)$, it is clear that $D_1(w_n)$ cannot exceed the length of the image of the segment joining the points $\zeta_n$ and $z_n$, since the point $\zeta_n$ is mapped onto a branch point of the Riemann surface. The length of this image is given by the integral

$$\int_{\zeta_n}^{z_n} |f'(z)|\,|dz| = \int_{\zeta_n}^{z_n} \left|\frac{2\sin 2W}{(1-z)^2}\right| |dz| \leqq \frac{2}{(1-z_n)^2}(z_n - \zeta_n)$$

$$= \frac{(1/n + 2\pi n + 1)^2}{n(1/n + 2\pi n + 1)(2n\pi + 1)}\cdot$$

Hence,

$$D_1(w_n) \leqq \frac{1/n + 2\pi n + 1}{n(2\pi n + 1)}, \qquad \lim_{n\to\infty} D_1(w_n) = 0.$$

Finally $w_n=\sin^2(1/n+2\pi n)=\sin^2 1/n$, so that $\lim_{n\to\infty} w_n=0$. This completes the proof of the theorem.

The idea of this example, as well as of the examples of §§12 and 23 is the

following. It is not true, as is well known[65], that $f_n(z)$ analytic for $|z|<1$, $f_n'(0)=1$, $f_n(0)=0$, implies that $w=f_n(z)$ maps $|z|<1$ onto a Riemann configuration which contains in its interior a fixed smooth circle whose center is at the origin. The simplest counterexample is perhaps

$$f_n(z) = z - nz^2.$$

The derivative $f_n'(z)=1-2nz$ vanishes for $z=1/2n$ and the corresponding value of $w$ is $f_n(1/2n)=1/4n$, which approaches zero.

This example indicates that the phenomenon of a branch point's approaching the origin is not dependent on the transcendentality of $f_n(z)$, or even on the possibility that an ever-increasing number of sheets of the image of $|z|<1$ should come together. It is a matter primarily of having the image of a point at which $f_n'(z)$ vanishes approach the origin. The examples mentioned above were constructed with this idea in mind.

## Chapter V. Miscellaneous

27. **Limit values of analytic functions.** The methods developed in the present paper have close connections with the general subject of limit values of functions analytic in the unit circle, including various theorems due to Lindelöf and to Montel. We proceed now to discuss such connections.

Theorem 1. *Let the function $f(z)$ be analytic for $|z|<1$ and omit two values there. Suppose for the sequence $\{z_n\}$ with $|z_n|<1$ we have $\lim_{n\to\infty} f(z_n)=\alpha$, where $\alpha$ is finite or infinite. Let the non-euclidean distance $\rho(z_n, z_n')$ between $z_n$ and $z_n'$ approach zero as $n$ becomes infinite, with $|z_n'|<1$. Then we have $\lim_{n\to\infty} f(z_n')=\alpha$.*

We define as usual the functions $g_n(\zeta)$:

$$g_n(\zeta) = f\left(\frac{\zeta + z_n}{1 + \bar{z}_n\zeta}\right), \tag{27.1}$$

whence $g_n(0)=f(z_n)$. If we set

$$z_n' = \frac{\zeta_n' + z_n}{1 + \bar{z}_n\zeta_n'}, \tag{27.2}$$

we have $g_n(\zeta_n')=f(z_n')$, and the non-euclidean distance

$$\rho(0, \zeta_n') = \rho(z_n, z_n') \tag{27.3}$$

approaches zero as $n$ becomes infinite. The family $g_n(\zeta)$ omits two values in $|\zeta|<1$, hence is normal there. Given any infinite sequence of indices $n$, there can be extracted a subsequence for which the corresponding functions $g_n(\zeta)$ converge for $|\zeta|<1$, uniformly in every closed subregion, to some limit func-

(65) See, for example, P. Montel, *Leçons sur les Fonctions Univalentes ou Multivalentes*, Paris, 1933, p. 121, where a different example is given.

tion $g(\zeta)$, with $g(0)=\lim_{n\to\infty} g_n(0)=\alpha$. The approach of zero to $\rho(0, \zeta_n')$ implies the approach to zero of $\zeta_n'$; so for the subsequence of indices considered the uniformity of convergence yields $\lim_{n\to\infty} g_n(\zeta_n')=\alpha$. Thus from any subsequence of the sequence $\{f(z_n')\}$ can be extracted a new subsequence converging to the limit $\alpha$, which implies the conclusion of Theorem 1.

THEOREM 2. *Let $f(z)$ be analytic for $|z|<1$ and omit two values there. Let the sequence $\{z_n\}$ with $|z_n|<1$ have the property that $\lim_{n\to\infty} f(z_n)=\alpha$, where $\alpha$ is finite. Then a necessary and sufficient condition that the sequence $\{z_n\}$ be regular*[66] *is*

$$\lim_{n\to\infty} g_n(\zeta) = \alpha \qquad \text{for } |\zeta| < 1, \tag{27.4}$$

*uniformly in every closed subregion, where $g_n(\zeta)$ is defined by* (27.1).

Let a sequence $\{z_n'\}$ be given for which $\rho(z_n, z_n')$ is bounded.

Again we define by $\zeta_n'$ equation (27.2), from which it follows that (27.3) is valid, and the non-euclidean distance $\rho(0, \zeta_n')$ is bounded. The sufficiency of (27.4) is obvious, for (27.4) implies that $g_n(\zeta_n')\to\alpha$, which is the conclusion to be established; we note that here the $\lambda$ of §11, Definition 1, can be taken arbitrarily large. We proceed to show the necessity of (27.4).

If the sequence $\{z_n\}$ is regular but (27.4) is not satisfied, there exists a sequence of indices $n_k$ such that $\lim_{k\to\infty} g_{n_k}(\zeta)=g_0(\zeta)$ for $|\zeta|<1$, uniformly in every closed subregion, where $g_0(\zeta)$ is analytic but not identically equal to $\alpha$ in $|\zeta|<1$. Suppose for definiteness $g_0(\zeta_0)\neq\alpha$, where the non-euclidean distance $\rho(0, \zeta_0)$ is less than the $\lambda$ of §11, Definition 1. If we define $z_n'$ by the equation

$$z_n' = \frac{\zeta_0 + z_n}{1 + \bar{z}_n\zeta_0},$$

we have

$$f(z_{n_k}') = g_{n_k}(\zeta_0) \to g_0(\zeta_0) \neq \alpha, \quad \rho(z_n, z_n') = \rho(0, \zeta_0) < \lambda,$$

contrary to hypothesis.

In Theorem 2 we have for simplicity assumed that $f(z)$ omits two values in $|z|<1$. It is obviously sufficient if $f(z)$ omits two values in the non-euclidean circle with non-euclidean center $z_n$ and non-euclidean radius $\rho_n$, where $\rho_n$ has a positive lower bound as $n$ becomes infinite. A similar remark applies to the later results of the present section.

A consequence of the foregoing remark is that if $f(z)$ is analytic for $|z|<1$, if $|z_n|<1$, if $\lim_{n\to\infty} f(z_n)=\alpha$, where $\alpha$ is finite, and if the sequence $\{z_n\}$ is irregular, then $f(z)$ has at most one omitted value in each set of non-euclidean circles with non-euclidean radius $\rho_n$, where $\rho_n$ has a positive lower bound. We

(66) For the definition of regularity see Definition 1 of §11, Chapter II.

consider pathology in more detail in §30. In Theorem 2, we have assumed the finiteness of $\alpha$. A result without this restriction appears in

COROLLARY 1. *Let $f(z)$ be analytic for $|z|<1$ and omit two values there. Let the sequence $\{z_n\}$ with $|z_n|<1$ have the property that $\lim_{n\to\infty} f(z_n)=\infty$. Then if $|z_n'|<1$ and if the non-euclidean distance $\rho(z_n, z_n')$ is bounded, we have also $\lim_{n\to\infty} f(z_n')=\infty$.*

Since the functions $g_n(\zeta)$ form a normal family in $|\zeta|<1$ and since $g_n(0)\to\infty$, we have $\lim_{n\to\infty} g_n(\zeta)=\infty$ in $|\zeta|<1$, uniformly in every closed subregion. Our conclusion is an immediate consequence. We turn to another result.

COROLLARY 2. *Let $f(z)$ be analytic for $|z|<1$ and omit there two values including the value $\alpha$. Let the sequence $\{z_n\}$ with $|z_n|<1$ have the property that $\lim_{n\to\infty} f(z_n)=\alpha$. Then if $|z_n'|<1$ and if the non-euclidean distance $\rho(z_n, z_n')$ is bounded, we have also $\lim_{n\to\infty} f(z_n')=\alpha$*[67].

From every infinite subsequence of the set $g_n(\zeta)$ defined by (27.1) can be extracted a new subsequence which converges for $|\zeta|<1$, uniformly in every closed subregion. The limit of this new subsequence is $\alpha$ in the point $\zeta=0$, hence by Hurwitz's theorem is identically $\alpha$ in $|\zeta|<1$. Then we have $\lim_{n\to\infty} g_n(\zeta)=\alpha$ for $|\zeta|<1$, uniformly in every subregion. Our conclusion follows as in the first part of the proof of Theorem 2. Thus the sequence $\{z_n\}$ is regular, and the number $\lambda$ of §11, Definition 1, may be chosen arbitrarily. A generalization of Theorem 2 is

COROLLARY 3. *Let $f(z)$ be analytic for $|z|<1$ and omit two values there, and let the sequence $w_n=f(z_n)$ with $|z_n|<1$ be bounded. Then a necessary and sufficient condition that the sequence $\{z_n\}$ be regular is*

$$\lim_{n\to\infty} [g_n(\zeta)-g_n(0)]=0 \qquad \textit{for } |\zeta|<1, \tag{27.5}$$

*uniformly in every closed subregion, where $g_n(\zeta)$ is defined by* (27.1).

If the sequence $\{z_n\}$ is regular, it follows that from any subsequence of the $g_n(\zeta)$ can be extracted a subsequence such that $\lim_{n\to\infty} g_n(\zeta)$ exists for $|\zeta|<1$, uniformly in every closed subregion; for this subsequence $\lim_{n\to\infty} g_n(0)=\alpha$ exists and by Theorem 2 the relation (27.5) holds for that subsequence. Thus, from any subsequence of the $g_n(\zeta)$ can be extracted a new subsequence such that (27.5) holds for that subsequence; so (27.5) itself is satisfied.

Conversely, if (27.5) is satisfied, and if $\rho(z_n, z_n')=\rho(0, \zeta_n')$ is bounded, it follows that $\lim_{n\to\infty} [g_n(\zeta_n')-g_n(0)]=0$, so the sequence $\{z_n\}$ is regular. Of

(67) This result for bounded functions was established by a different method by one of the present authors: W. Seidel, these Transaction, vol. 34 (1932), pp. 1–21; especially Theorem 3, p. 10.

course, this latter conclusion is independent of any assumption that $f(z)$ omit two values.

Two further propositions relate Theorem 2 to the results of §§17 and 22.

COROLLARY 4. *Let the function $f(z)$ be analytic in $|z|<1$ and omit two values there, and let the sequence $\{f(z_n)\}$ be bounded, $|z_n|<1$. A necessary and sufficient condition that $\{z_n\}$ be a regular sequence for $f(z)$ is*

$$\lim_{n\to\infty} |f^{(k)}(z_n)|(1-|z_n|)^k = 0, \qquad k = 1, 2, 3, \cdots. \tag{27.6}$$

From the sequence $f(z_n)$ can be extracted a subsequence $f(z_{n_j})$ which approaches a limit $\alpha$. A necessary and sufficient condition for (27.4) for the sequence $\{n_j\}$ is

$$\lim_{n_j\to\infty} g_{n_j}^{(k)}(0) = 0, \qquad k = 1, 2, 3, \cdots, \tag{27.7}$$

since the functions $g_n(\zeta)$ form a normal family in $|\zeta|<1$. Equations (27.6) are equivalent to equations (27.7) if the latter are assumed to hold for a suitable subsequence $\{n_j\}$ of an arbitrary sequence of indices.

COROLLARY 5. *Let the function $f(z)$ be analytic and omit two values in $|z|<1$. A necessary and sufficient condition that $\{z_n\}$ be a regular sequence for $f(z)$, where we assume $w_n=f(z_n)$ bounded, is*

$$\lim_{n\to\infty} D_p(w_n) = 0, \qquad p = 1, 2, 3, \cdots.$$

Corollary 5 follows from Corollary 4 by virtue of our fundamental Theorem 1 of §22.

A further consequence of Corollary 5 is

COROLLARY 6. *Let the function $f(z)$ be analytic in $|z|<1$ and omit two values there. If the sequence of points $w_n=f(z_n)$, with $|z_n|<1$, approaches a finite boundary point of the Riemann configuration on which $w=f(z)$ maps $|z|<1$, then the sequence $\{z_n\}$ is regular.*

It is worth remarking that Corollary 2 is a consequence of Corollary 5 or Corollary 6, without the use of Hurwitz's theorem.

Theorems 1 and 2 are of particular interest if the function $f(z)$ approaches a limit along an arc.

THEOREM 3. *Let $f(z)$ be analytic in $|z|<1$ and omit two values there. Let the Jordan arc $C$ lie in $|z|<1$ except for the end point $z=1$. Suppose*

$$\lim_{z\to 1,\, z \text{ on } C} f(z) = \alpha, \tag{27.8}$$

*where $\alpha$ is finite. Then any sequence $\{z_n\}$ on $C$ for which $z_n\to 1$ is regular.*

From any subsequence of the sequence $g_n(\zeta)$ defined by (27.1) can be extracted a new subsequence converging to some function $g_0(\zeta)$ for $|\zeta|<1$, uniformly for $|\zeta|\leqq d<1$. Let $h$ be arbitrary, $0<h<\infty$. Let $z_n'$ be a point of $C$ between the points $z_n$ and $z=1$ with $\rho(z_n, z_n')=h$; such a point $z_n'$ exists with $\lim_{n\to\infty} z_n'=1$. If $\zeta_n'$ is defined by (27.2), we have $\rho(0, \zeta_n')=h$. The equation (27.8) implies $\lim_{n\to\infty} f(z_n')=\alpha$, whence $\lim_{n\to\infty} g_n(\zeta_n')=\alpha$. On each circle $|\zeta|=d<1$ lies a sequence of points $\zeta_n'$ for which $g_n(\zeta_n')$ approaches $\alpha$, so on each such circle lies at least one point $\zeta$ at which $g_0(\zeta)=\alpha$. Consequently $g_0(\zeta)\equiv\alpha$ in $|\zeta|<1$, every limit function of the sequence $g_n(\zeta)$ is identically $\alpha$, this limit is approached by $g_n(\zeta)$ itself throughout $|\zeta|<1$, uniformly in any closed subregion; our conclusion follows from Theorem 2.

The method of proof of Theorem 3 establishes also the following: *Let $f(z)$ be analytic in $|z|<1$ and omit two values. Let $z_n\to 1$, with $|z_n|<1$, and let the non-euclidean distance $\rho(z_n, z_{n+1})$ approach zero. If* $\lim_{n\to\infty} f(z_n)$ *exists, then the sequence $\{z_n\}$ is regular.* By way of proof, we need merely modify the proof of Theorem 3 by considering instead of the arbitrary circle $|\zeta|=d<1$ an arbitrary annulus $0<d_1\leqq|\zeta|\leqq d_2<1$; each such annulus contains a sequence of points $\zeta_n'$ for which $g_n(\zeta_n')$ approaches $\alpha$, so each closed annulus contains at least one point $\zeta$ in which $g_0(\zeta)=\alpha$.

The method of proof of Theorem 3 can be used to prove still another proposition: *Let $f(z)$ be analytic in $|z|<1$ and omit there two values. Suppose for real $z$ we have* $\lim_{z\to 1} f(z)=\alpha$. *Then we have uniformly for approach within any triangle in $|z|<1$*

$$\lim_{z\to 1} f^{(k)}(z)(1-|z|^2)^k = 0.$$

In this proof, we need merely choose $r$, $0<r<1$, and the sequence of real $z_n$ in such a way that under the transformation $z=(\zeta+z_n)/(1+\bar{z}_n\zeta)$ each point $z$ of the given triangle corresponds to some $\zeta$ in $|\zeta|<r$. Various extensions of the proposition by the present methods suggest themselves, and are left to the reader.

We shall introduce the notion of the ***non-euclidean Fréchet distance between two curves***. Let $C_1$ and $C_2$ be two open Jordan arcs lying in $|z|<1$. Consider a topological map $T$ of $C_1$ on $C_2$. Denote by $F_T(C_1, C_2)$ the least upper bound (finite or infinite) of the non-euclidean distances between points of $C_1$ and $C_2$ which correspond in the map $T$. The greatest lower bound (finite or infinite) of the quantities $F_T(C_1, C_2)$ for all possible maps $T$ will be called the ***non-euclidean Fréchet distance $F(C_1, C_2)$ between $C_1$ and $C_2$***. With this definition we prove

THEOREM 4. *Let $f(z)$ be analytic in $|z|<1$ and omit two values there. Let $C_1$ and $C_2$ be Jordan arcs which, except for the common end point $z=1$, lie in $|z|<1$, and let $F(C_1, C_2)$ be finite. If*

$$\lim_{z\to 1;\, z \text{ on } C_1} f(z) = \alpha,$$

*where $\alpha$ is finite or infinite, then also*

$$\lim_{z\to 1;\, z \text{ on } C_2} f(z) = \alpha.$$

To any sequence $z_n'$ on $C_2$ which approaches $z=1$ corresponds a sequence $z_n$ on $C_1$ such that $\rho(z_n, z_n')$ is bounded. If $\alpha$ is finite, our conclusion follows from Theorem 3. If $\alpha$ is infinite, it follows from Corollary 1 to Theorem 2.

If the two Jordan arcs $C_1$ and $C_2$ of Theorem 4 are tangent and have the same order of contact with $|z|=1$ at $z=1$, then $F(C_1, C_2)$ is finite. For transform by a linear transformation of the complex variable the region $|z|<1$ onto the upper half of the $w$ $(=x+iy)$-plane, so that $z=1$ corresponds to $w=0$. We shall assume that in the neighborhood of $w=0$ we may set up a one-to-one correspondence between the arcs $C_1: y=y_1(x)$ and $C_2: y=y_2(x)$ by means of the ordinates $x=\text{constant}$. The non-euclidean distance between corresponding points of the two curves reduces to

$$\left|\log\frac{y_1(x)}{y_2(x)}\right|,$$

which by the assumption on order of contact is bounded. In studying the finiteness of $F(C_1, C_2)$, we may confine ourselves to the neighborhood of the point $z=1$, so under the present hypothesis $F(C_1, C_2)$ is finite.

If the two Jordan arcs $C_1$ and $C_2$ of Theorem 4 are, except for the point $z=1$, contained in the lens-shaped region between two hypercycles through $z=\pm 1$, and possess tangents at the point $z=1$, we may set up a one-to-one correspondence between their points by the circles of the coaxial family determined by $z=\pm 1$ as null circles. Transformation of an arbitrary circle of that family into the axis of imaginaries by a transformation which leaves invariant $z=1$, $z=-1$, and $|z|=1$, as well as the two given hypercycles, shows that $F(C_1, C_2)$ is finite. Thus we have the

COROLLARY. *The condition of Theorem 4 that $F(C_1, C_2)$ be finite is satisfied if $C_1$ and $C_2$ are tangent and have contact of the same order with $|z|=1$ at $z=1$, or if $C_1$ and $C_2$ possess tangents at $z=1$ but neither is tangent to $|z|=1$ at $z=1$.*

The foregoing discussion has intimate connections with well known results on the limit values of analytic functions. The proof of Theorem 2 establishes the uniformity for all $z_n'$ of $\lim_{n\to\infty} f(z_n')$ provided merely $\rho(z_n, z_n')$ is uniformly bounded. With this addition, Theorem 4 and its corollary include the theorem of Lindelöf that if $f(z)$ is analytic in $|z|<1$ and omits two values there, and if $\lim_{z\to 1} f(z)$ exists for approach along a line segment in $|z|\leq 1$, then that limit exists uniformly for approach within an arbitrary triangle contained in

$|z| \leqq 1$. Likewise the corollary to Theorem 4 includes the theorem of Montel that if $f(z)$ is analytic in $|z| < 1$ and omits two values there, and if $\lim_{z\to 1} f(z)$ exists for approach along the arc of an oricycle, then that limit exists uniformly for approach between any two arcs of oricycles tangent at $z=1$ to the original arc.

We add the general remark that the method of the present section seems to have further wide use in the study of limit values of analytic functions; for instance this method easily proves that *if $f(z)$ is analytic and bounded in $|z| < 1$, continuous on $|z| = 1$ or an open arc $A$ of $|z| = 1$ with $z=1$ as an end point, and if on this arc* $\lim_{z\to 1,\ z \text{ on } A} f(z) = \alpha$, *then also the limit of $f(z)$ is $\alpha$ uniformly as $|z| \to 1$ between $A$ and the axis of reals.*

28. **Extension of Bloch's theorem.** Another application of the results of Chapter III deals with an extension of Bloch's theorem. We prove the following, which for $p=1$ reduces to Bloch's theorem.

THEOREM 5. *Let $w=f(z)$ be regular in $|z| < 1$ and let $f^{(p)}(0) = 1$. There exists an absolute positive constant $B_p$, independent of the function $f(z)$ so that the Riemann configuration $R_f$ on which $w=f(z)$ maps the circle $|z| < 1$ contains at least one point $w_0$ for which $D_p(w_0) \geqq B_p$. The constant $B_p$ may be taken equal to $2^{-p}\lambda_p$, where $\lambda_p = M_p/p!$, $M_p = M_p(M)$ being the constant of Theorem* 1, *Chapter* III, *taken for $M = 2^p \cdot p!$.*

We assume first that $f(0) = 0$ and that $f(z)$ is regular in $|z| \leqq 1$. Let

$$M_p(r) = \max_{|z| \leqq r} |f^{(p)}(z)|.$$

We have $M_p(0) = 1$ and the function $M_p(r)$ is continuous and non-decreasing in the interval $0 \leqq r \leqq 1$. The function

$$\phi(r) = (1 - r)^p M_p(r)$$

is also continuous in $0 \leqq r \leqq 1$ and $\phi(0) = 1$, $\phi(1) = 0$. Hence, there exists a number $r_0$ $(0 \leqq r_0 < 1)$ such that $\phi(r_0) = 1$ and $\phi(r) < 1$ for $r_0 < r \leqq 1$. The function $|f^{(p)}(z)|$ attains the value $M_p(r_0)$ at a point $z_0$ of modulus $r_0$:

$$(28.1) \qquad |f^{(p)}(z_0)| = M_p(r_0) = \frac{1}{(1 - r_0)^p}.$$

Consider a circle $\gamma$ of center $z_0$ and radius $\rho = (1 - r_0)/2$ and the function

$$g(\zeta) = \frac{f(z_0 + \rho\zeta) - f(z_0)}{\rho^p f^{(p)}(z_0)} = a_1\zeta + a_2\zeta^2 + \cdots + \frac{\zeta^p}{p!} + \cdots$$

for suitably chosen constants $a_1, a_2, \cdots$. It is regular in $|\zeta| \leqq 1$ and

$$g^{(p)}(\zeta) = \frac{f^{(p)}(z_0 + \rho\zeta)}{f^{(p)}(z_0)}.$$

Now, in the circle $|\zeta| \leqq 1$ we have $|z_0+\rho\zeta| \leqq r_0+(1/2)(1-r_0)=(1/2)(1+r_0)$ and therefore in $|\zeta| \leqq 1$

$$|f^{(p)}(z_0+\rho\zeta)| \leqq M_p\left(\frac{1+r_0}{2}\right) < \frac{1}{(1-(1/2)(1+r_0))^p} = \frac{2^p}{(1-r_0)^p}.$$

Hence, in view of (28.1)

$$|g^{(p)}(\zeta)| < 2^p$$

for $|\zeta| \leqq 1$. Successive integration shows that

$$|g(\zeta)| < 2^p \tag{28.2}$$

for $|\zeta| \leqq 1$ and we also have

$$g(0) = 0, \qquad g^{(p)}(0) = 1. \tag{28.3}$$

Now it was shown in Theorem 1, Chapter III, that for the class of functions satisfying the conditions (28.2) and (28.3)

$$D_p(0) \geqq \frac{M_p(2^p \cdot p!)}{p!} = \lambda_p. \tag{28.4}$$

Consequently, by the definition of $g(\zeta)$ it follows that for the function $f(z)$, setting $w_0=f(z_0)$,

$$D_p(w_0) \geqq \lambda_p \cdot \rho^p |f^{(p)}(z_0)| = \lambda_p/2^p.$$

The condition that $f(z)$ be analytic in the closed circle $|z| \leqq 1$ may now be lifted. Indeed, let $f(z)$ be assumed to be analytic in $|z| < 1$. Then, if $r$ is a value in the interval $0<r<1$, the function

$$F(z) = \frac{1}{r^p} f(rz)$$

is analytic for $|z| \leqq 1$ with $F^{(p)}(0)=1$. Furthermore, we have

$$D_p[F(z)] \leqq \frac{1}{r^p} D_p[f(rz)].$$

Hence, since the theorem applies to $F(z)$, there exists a $z_0$ ($|z_0| < 1$) so that

$$B_p \leqq \frac{1}{r^p} D_p[f(rz_0)].$$

Now allowing $r$ to approach one, we obtain the theorem in the general case.

A lower bound for $B_p$ may be obtained from the estimate in (19.2). This value, however, is certainly not sharp.

As a matter of record, we formulate without proof the

COROLLARY. *Let the function $w=f(z)$ be analytic in $|z|<1$, with*

$$|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)| = m.$$

*There exists a positive constant $B_p'$ independent of $m$ and $f(z)$ such that the Riemann configuration $R_f$ onto which $w=f(z)$ maps the region $|z|<1$ contains at least one point for which $D_p(w_0) \geqq mB_p'$. In fact, we may choose $B_p'$ as the smallest of the numbers $j!\cdot B_j/p$, $j=1, 2, \cdots, p$, in the notation of Theorem 5.*

29. **Unrestricted functions; properties of $\Delta(z)$.** From the example of §26 it is clear at once that one cannot obtain a relation between $D_1(w_0)$ and $|f'(z_0)|(1-|z_0|^2)$ without some restriction on the class of functions $f(z)$ to be considered. It is perhaps not without interest to remark that by introducing a new quantity $\Delta(z_0)$ one may obtain relations of the desired kind without imposing any restriction on $f(z)$ other than analyticity in the unit circle $|z|<1$. In fact, we prove

THEOREM 6. *Let $w=f(z)$ analytic for $|z|<1$ map $|z|<1$ onto a Riemann configuration $S$. Let $z_0$ be any point of the circle $|z|<1$ which is mapped by $w=f(z)$ onto a point $w_0$ of $S$ which is not a branch point of $S$. Denoting by $\Delta(z_0)$ the radius of the largest circle of the $\zeta$-plane with center $\zeta=0$ in which the function*

$$\phi(\zeta) = f\left(\frac{\zeta + z_0}{1 + \bar{z}_0\zeta}\right)$$

*is univalent, the inequality*

$$\frac{1}{4}\,\frac{D_1(w_0)}{\Delta(z_0)} \leqq |f'(z_0)|\,(1 - |z_0|^2) \leqq 4\,\frac{D_1(w_0)}{\Delta(z_0)} \tag{29.1}$$

*holds. In particular, for a sequence of points $\{z_n\}$ ($|z_n|<1$) a necessary and sufficient condition for*

$$\lim_{n\to\infty} |f'(z_n)|\,(1 - |z_n|^2) = 0 \tag{29.2}$$

*is*

$$\lim_{n\to\infty} \frac{D_1(w_n)}{\Delta(z_n)} = 0, \tag{29.3}$$

*where $w_n=f(z_n)$.*

The function $\zeta=\psi(w)$ inverse to $w=\phi(\zeta)$ is univalent on $S$, in particular univalent for $|w-w_0|<D_1(w_0)$. Therefore, by Koebe's distortion theorem it must map the circle $|w-w_0|<D_1(w_0)$ onto some region of the $\zeta$-plane within

which $\phi(\zeta)$ is univalent and which contains in its interior the circle

$$|\zeta| < (1/4)|\psi'(w_0)|\cdot D_1(w_0),$$

whence

$$\Delta(z_0) \geqq (1/4)|\psi'(w_0)|\cdot D_1(w_0).$$

By the relation $1/\psi'(w_0)=\phi'(0)=f'(z_0)\ (1-|z_0|^2)$ we obtain the left-hand side of inequality (29.1).

Similarly, the function $w=\phi(z)$ is univalent for $|\zeta|<\Delta(z_0)$, hence again by Koebe's distortion theorem maps smoothly $|\zeta|<\Delta(z_0)$ onto a region containing the circle $|w-w_0|<(1/4)\ |\phi'(0)|\cdot\Delta(z_0)$. Hence,

$$D_1(w_0) \geqq (1/4)|\phi'(0)|\cdot\Delta(z_0),$$

and the right-hand side of inequality (29.1) follows directly.

Next, the equivalence of the relations (29.2) and (29.3) follows from (29.1) provided $w_n$ are not branch points of $S$. Indeed, if $w_0$ is a branch point of $S$, the expression $D_1(w_0)/\Delta(z_0)$ has no meaning since both numerator and denominator are zero. We observe, however, that if a sequence of points $z_n$ for which the corresponding points $w_n$ are not branch points of $S$ converge to a point $z_0$ (with $|z_0|<1$) for which the corresponding point $w_0$ is a branch point of $S$, then by the first inequality of (29.1)

$$\lim_{n\to\infty}\frac{D_1(w_n)}{\Delta(z_n)} = 0.$$

Hence, it is reasonable to define $D_1(w_0)/\Delta(z_0)$ as zero when $w_0$ is a branch point of $S$. With this convention the equivalence of (29.2) and (29.3), as well as the inequality (29.1), remain valid even in the case of branch points.

30. **Pathology.** There are several fairly obvious extensions of our fundamental Theorem 2 of Chapter IV to the effect that if $f(z)$ is analytic and omits two values in $|z|<1$, if $\{z_n\}$ is a sequence of points in $|z|<1$, and if the numbers $w_n=f(z_n)$ are bounded, then the two conditions

$$D_1(w_n) \to 0, \tag{30.1}$$

$$f'(z_n)(1-|z_n|^2) \to 0, \tag{30.2}$$

are equivalent in the sense that each implies the other. The mere analyticity of $f(z)$ insures that (30.2) implies (30.1); so we are concerned at present only with the condition that (30.1) shall imply (30.2). Thus it is sufficient for (30.1) to imply (30.2) if we replace the condition that $f(z)$ omits two values in $|z|<1$ by the condition that

$$\phi_n(\zeta) = f\left(\frac{\zeta+z_n}{1+\bar{z}_n\zeta}\right) \tag{30.3}$$

shall omit two values in $|\zeta| < r < 1$, where $r$ is independent of $n$; no essential change in the original reasoning is necessary; compare §27, Corollaries 4 and 5. It is obvious too that (30.1) implies (30.2) provided from each subsequence $z_{n_k}$ of the $z_n$ can be extracted a new subsequence $z_{m_k}$ for which there exists a positive number $r<1$ such that the corresponding functions $\phi_{m_k}(\zeta)$ defined by (30.3) have two omitted values in $|\zeta| < r$; for under such circumstances the fulfillment of condition (30.1) implies that for no subsequence $z_{n_k}$ does the expression

$$f'(z_{n_k})(1 - |z_{n_k}|)$$

approach a limit different from zero, whence (30.2) is satisfied. For instance, it may occur that the functions $\phi_{2n}(\zeta)$ have the exceptional values 0 and 1 in $|\zeta| < 1/2$, and that the functions $\phi_{2n+1}(\zeta)$ have the exceptional values 2 and 3 in $|\zeta| < 1/4$.

DEFINITION. *Let the function $f(z)$ be analytic for $|z| < 1$, let $\{z_n\}$ be a sequence of points in $|z| < 1$, let $w_n = f(z_n)$ approach a finite limit, let* (30.1) *be satisfied but suppose*

$$\lim_{n\to\infty} f'(z_n)(1 - |z_n|) = \alpha \neq 0; \tag{30.4}$$

*then we shall say that $\{z_n\}$ is a q-sequence.*

The discussion we have already given yields

THEOREM 7. *Under the hypothesis of the italicized definition, let $\{z_n\}$ be a q-sequence. Then from no subsequence $\{z_{n_k}\}$ of the $\{z_n\}$ can there be extracted a new subsequence $\{z_{m_k}\}$ such that the functions $\phi_{m_k}(\zeta)$ defined by* (30.3) *have two exceptional values in any region $|\zeta| < r < 1$, where $r$ is independent of $m_k$.*

In other words, if $\{z_n\}$ is a $q$-sequence, then for every $r$, $0<r<1$, and for every infinite sequence of subscripts $\{n_k\}$, the functions $\phi_{n_k}(\zeta)$ have at most one exceptional value in $|\zeta| < r$.

Some consequences of Theorem 7 are more conveniently described after transformation of $|z| < 1$ onto a half-plane $\Re(z') > 0$.

THEOREM 8. *Under the hypothesis of the italicized definition, let $\{z_n\}$ be a q-sequence having as limit the point $z_0$, with $|z_0| = 1$. Let the region $|z| < 1$ be transformed by a linear transformation onto $\Re(z') > 0$ so that $z = z_0$ corresponds to $z' = 0$. Then there exists a half-line $L$ from $z' = 0$ in the closed region $\Re(z') \geqq 0$ possessing the property that if $S$ is a sector (of a circle) containing $L$ in its interior and with vertex in $z' = 0$, of arbitrarily small radius, then in $S$ the transform of the function $f(z)$ has at most one exceptional value.*

Let the points $z_n'$ (necessarily approaching $z' = 0$) be the transforms in the

$z'$-plane of the points $z_n$. The numbers

$$\theta_n = \arg z_n', \qquad -\pi < \theta n < \pi$$

have at least one limit value, say $\theta = \theta_0$; *the half-line $L$ may be chosen as $\theta = \theta_0$, as we shall proceed to prove.*

A non-euclidean circle in the $z'$-plane whose non-euclidean center is $z' = \alpha$, $\Re(\alpha) > 0$, is transformed by shrinking or stretching the plane with $z' = 0$ fixed into a non-euclidean circle of the same radius, for the transformation leaves the region $\Re(z') > 0$ invariant. Let $S$ be given, and let $S'$ be a sector interior to $S$ whose sides are also interior to $S$, likewise having $z' = 0$ as vertex, and containing $L$ in its interior. Then an infinity of points $z_n'$ lie interior to $S'$. Let $\rho$ denote the smaller of the two non-euclidean radii of the two circles whose euclidean centers lie on the respective rays bounding $S'$ and which are tangent to $S$; the circles are not uniquely determined but their non-euclidean radii are uniquely determined; there is an exceptional situation here, which presents no inherent difficulty and whose treatment is left to the reader, if the half-line $\theta = \pi/2$ or $\theta = -\pi/2$ lies in or on the boundary of $S$. The non-euclidean circles whose common non-euclidean radius is $\rho$ and whose euclidean centers are the infinity of points $z_n'$ interior to $S'$ all of whose interior points are interior points of $S$. Theorem 8 now follows from Theorem 7.

It is obviously true that in $S$ the function $f(z)$ takes on every value with at most one exception *an infinite number of times.*

Theorem 8 obviously bears a close analogy to Julia's theorems on entire functions[48]. The analogy can be pursued still more closely as we now indicate.

In the $z'$-plane used in Theorem 8 let $C$ be an arbitrary curve (not necessarily a Jordan curve) joining the unit circle to the origin.

$$C:\quad z' = \sigma(t), \qquad 0 \leqq t \leqq 1,$$
$$\sigma(0) = 0, \qquad |\sigma(1)| = 1,$$

where $\sigma(t)$ is a continuous complex-valued function of the real parameter $t$. From $C$ is found by rotation about the origin a curve which we denote by $C(\omega)$: $z' = \omega \cdot \sigma(t)$, $|\omega| = 1$. We shall call a *horn* the set $H(\omega, \epsilon)$ of points each of which lies interior to at least one of the circles having its center in a point $z'$ on $C(\omega)$ and of radius $\epsilon \cdot |z'|$. It will be noted that the horn $H(\omega, \epsilon)$ is then a region, and that each of its boundary points except $z' = 0$ is on the circumference of a circle of center $z'$ on $C(\omega)$ and radius $\epsilon|z'|$. But of course the curve $C$ and the horn $H(\omega, \epsilon)$ need not lie entirely in the closed region $\Re(z') \geqq 0$.

We now prove a generalization of Theorem 8.

THEOREM 9. *Under the hypothesis of Theorem 8 for arbitrary $C$ there exists*

[48] G. Julia, *Leçons sur les Fonctions Uniformes à Point Singulier Essentiel Isolé*, Paris, 1924, p. 105 ff.

*a curve $C(\omega_0)$ such that in every horn $H(\omega_0, \epsilon)$ the transform of the function $f(z)$ takes on every value an infinite number of times, with the exception of at most one value.*

Let the numbers $\epsilon$ and $\epsilon_1$ be given, $1>\epsilon>\epsilon_1>0$. Consider all circles $\gamma$ and $\gamma_1$ of radii $r\epsilon$ and $r\epsilon_1$ with variable common center $(r, \theta)$, where $r$ is bounded and $\theta$ is arbitrary. Then the non-euclidean distance from a point of $\gamma_1$ in the region $\Re(z')>0$ to the nearest point of $\gamma$ is bounded from zero, say is greater than or equal to some positive $\delta$ independent of $r$ and $\theta$. This conclusion follows from the fact that in studying the non-euclidean distance it is no loss of generality to take $r=1$.

As an application of this remark, since each boundary point of the horn $H(\omega, \epsilon_1)$, lies on a circle $\gamma_1$ with center $z'$ on $C(\omega)$ and radius $\epsilon_1|z'|$, and since all points interior to the circle $\gamma$ with center $z'$ and radius $\epsilon|z'|$ belong to $H(\omega, \epsilon)$, it follows that the non-euclidean distance from each boundary point of $H(\omega, \epsilon_1)$ in $\Re(z')>0$ to the boundary of $H(\omega, \epsilon)$ is greater than or equal to $\delta$. *If all points of a set $\{z'_{n_k}\}$ in $\Re(z')>0$ lie in $H(\omega, \epsilon_1)$, then each point whose non-euclidean distance from some $z_{n_k}$ is less than $\delta$ lies in $H(\omega, \epsilon)$.*

Suppose now the points

$$z_n' = r_n e^{i\theta_n}, \qquad 0 < r_n \leqq 1; n = 1, 2, \cdots ,$$

are the transforms in the $z'$-plane of the given $q$-sequence. Each $z_n'$ lies on some curve $C(\omega_n)$; in fact, the continuous function $|\sigma(t)|$ must take on the value $r_n$ for some value of $t$, say $t_n$, $0<t_n\leqq 1$, whence

$$z_n' = | \sigma(t_n) | e^{i\theta_n}, \qquad\qquad r_n = | \sigma(t_n) |,$$

so $z_n'$ lies on the curve

$$C(\omega_n):\ \ z' = \omega_n \cdot \sigma(t), \qquad \omega_n = \frac{e^{i\theta_n} | \sigma(t_n) |}{\sigma(t_n)} \cdot$$

Of course $t_n$ and $\omega_n$ need not be uniquely defined, but we choose a specific determination.

Let the set $\omega_1, \omega_2, \cdots$ on the unit circle have the limit point $\omega_0$. *Then for every $\epsilon_1>0$, the horn $H(\omega_0, \epsilon_1)$ has an infinity of the points $z_n'$ in its interior.* For on the arc of $|z'|=1$ in the circle $|z'-\omega_0|=\epsilon_1$ lie an infinity of points $\omega_n$, say $\omega_{n_1}, \omega_{n_2}, \cdots$. Then of the circle $r=r_{n_k}$ the entire arc which lies in the circle

$$| z' - \omega_0 \cdot \sigma(t_{n_k}) | = \epsilon_1 \cdot r_{n_k}$$

lies on $H(\omega_0, \epsilon_1)$, and this arc of the circle $r=r_{n_k}$ contains the point

$$z'_{n_k} = | \sigma(t_{n_k}) | e^{i\theta_{n_k}}$$

by virtue of the inequality for $\omega_{n_k}$

$$\left| \frac{e^{i\theta_{n_k}} |\sigma(t_{n_k})|}{\sigma(t_{n_k})} - \omega_0 \right| \leqq \epsilon_1.$$

We are now in a position to prove Theorem 9. Let $C$ be given. The number $\omega_0$ is to be determined as just indicated, and thus $C(\omega_0)$ is defined. Then $\epsilon > 0$ is arbitrary, and we choose $\epsilon_1$, $0 < \epsilon_1 < \epsilon$. The points $z'_{n_k}$ already defined are the transforms of a $q$-sequence $z_{n_k}$; it follows from Theorem 7 that in the set of circles having the $z'_{n_k}$ as non-euclidean centers and with a common non-euclidean radius the function $f_1(z')$ [transform of $f(z)$] takes on every value with at most one exception an infinite number of times. The points $z'_{n_k}$ lie in $H(\omega_0, \epsilon_1)$, and these non-euclidean circles (chosen with common non-euclidean radius less than the number $\delta$ previously defined) all lie in $H(\omega_0, \epsilon)$. The proof is complete.

It is also true that in every $H(\omega_0, \epsilon)$ *in every neighborhood of the origin* the function $f_1(z')$ takes on every value with at most one exception an infinite number of times.

31. **Functions with bounded $D_1$.** In studying the relation between $D_1(w)$ and $|f'(z)|(1-|z|^2)$ we have restricted the class of functions $f(z)$ in such a manner that the associated functions $\phi_n(\zeta)$ should form a normal family. For this reason we considered the class of univalent functions, the class of bounded functions, and the class of functions omitting two values. There is, however, another criterion of normality, which was discovered by Bloch([69]). It is the class of functions for which the radius of univalence $D_1(w)$ is bounded. The desired relations may be easily obtained for this class. Indeed, we have

THEOREM 10. *Let $w=f(z)$ be analytic for $|z|<1$, and let $D_1(w)$ be uniformly bounded: $D_1(w) \leqq D$. Setting $w_0 = f(z_0)$, where $z_0$ is an arbitrary point of $|z|<1$, the inequality*

$$|f'(z_0)|(1-|z_0|^2) \leqq [K \cdot D_1(w_0)]^{1/2}, \tag{31.1}$$

*holds, where $K$ may be taken equal to $20D/B$, $B$ being Bloch's constant.*

We begin by using the method of proof (Montel, ibid.) of Bloch's theorem on normal families. If we set

$$\phi(z) = f\left(\frac{z+z_0}{1+\bar{z}_0 z}\right), \qquad g(\zeta) = \frac{\phi[z_1 + (1-|z_1|)\zeta]}{(1-|z_1|)\phi'(z_1)}, \tag{31.2}$$

$$|z_0| < 1, \ |z_1| < 1, \ \phi'(z_1) \neq 0,$$

we note that $g(\zeta)$ is analytic in $|\zeta|<1$, with $g'(0)=1$. Then it follows from Bloch's theorem (§28, Theorem 5 for $p=1$) that for the function $g(\zeta)$ and for some $w$ we have $D_1(w) \geqq B$, where $B$ is Bloch's constant; hence if $D_1(w)$ refers now to the function $\phi[z_1+(1-|z_1|)\zeta]$ we have for some $w$

([69]) Cf. P. Montel, ibid., p. 115.

$$\frac{D_1(w)}{(1-|z_1|)\,|\phi'(z_1)|} \geqq B.$$

But by our hypothesis we have $D_1(w) \leqq D$, whence

$$|\phi'(z_1)| \leqq \frac{D}{B(1-|z_1|)};$$

this inequality is valid in the case $\phi'(z_1)=0$, exceptional for (31.2).

If we introduce the notation

$$\Phi(\zeta) = \phi(\zeta) - \phi(0) = \int_0^{\zeta} \phi'(\zeta)d\zeta, \tag{31.3}$$

where the integral is taken along a line segment, we have for $|\zeta| \leqq \rho < 1$

$$|\Phi(\zeta)| \leqq \int_0^{\rho} \frac{Dd\rho}{B(1-\rho)} = -\frac{D}{B}\log(1-\rho).$$

The inequality of Theorem 2, §10, can be written in the present case

$$D_1(w_0) \geqq \frac{|\phi'(0)|^2\rho^2}{-(4D/B)\log(1-\rho)}, \qquad w_0 = f(z_0).$$

It is seen immediately that the maximum of the function

$$-\frac{\rho^2}{\log(1-\rho)}, \qquad 0 < \rho < 1,$$

occurs when

$$-\log(1-\rho) = \frac{\rho}{2(1-\rho)}$$

which is approximately $\rho = .72$, so we may take

$$D_1(w_0) \geqq \frac{B}{10D}|f'(z_0)|^2(1-|z_0|^2)^2.$$

This proves the theorem.

As a corollary, it is seen that under the hypothesis of Theorem 10 the condition $D_1(w_n)\to 0$ is a sufficient condition for $|f'(z_n)|\,(1-|z_n|)\to 0$ even when $w_n\to\infty$. As a further remark it may be observed that the class of functions considered in Theorem 10 includes the case that the area of the image of $|z|<1$ under the transformation $w=f(z)$ is finite.

It is clear that analogous inequalities could be obtained for the higher derivatives. We proceed instead to the analogous theorem for $D_p(w)$ in general:

THEOREM 11. *Let $w=f(z)$ be analytic for $|z|<1$, and let $D_p(w)$ be uniformly bounded: $D_p(w)\leqq D_p$, where $p$ is given and $D_p$ is independent of $w$. If we set $w_0=f(z_0)$, where $|z_0|<1$, we have*

$$\sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^{\nu}C_{k-1,\nu}\bar{z}_0^{\nu}\frac{(1-|z_0|^2)^{k-\nu}f^{(k-\nu)}(z_0)}{(k-\nu)!}\right|$$

$$\leqq 24pK_p\left(\frac{D_p}{B_p'}\right)^{1-2^{-p}}[D_p(w_0)]^{2^{-p}},$$

*where $B_p'$ is the constant of the corollary of §28, and where $K_p$ is a constant depending only on $p$; indeed we may set*

$$K_p=\min\left\{\rho^{-p}[-\log(1-\rho)]^{1-2^{-p}},\ 0<\rho<1\right\},$$

*or we may set $K_p=2^p$.*

Of course the boundedness of $D_p(w)$, as in Theorem 11, is a stronger condition than the boundedness of $D_1(w)$, as in Theorem 10, for we have $D_p(w)\geqq D_1(w)$.

As before, we introduce $\phi(z)$ by the first of equations (31.2), but set now $G(\zeta)=\phi[z_1+(1-|z_1|)\zeta]$, where $z_1$ is arbitrary provided $|z_1|<1$. Thus $G(\zeta)$ is analytic in $|\zeta|<1$. Then if $D_p(w_0)$ refers to $G(\zeta)$ or to $f(z)$, we have by the corollary, §28

$$D_p(w_0)\geqq B_p'\left[|\Phi'(0)|+\frac{1}{2!}|\Phi''(0)|+\cdots+\frac{1}{p!}|\Phi^{(p)}(0)|\right].$$

By virtue of the inequality $D_p(w_0)\leqq D_p$, we may now write

$$\frac{D_p}{B_p'}\geqq(1-|z_1|)|\phi'(z_1)|.$$

In the notation of (31.3) we have for $|\zeta|\leqq\rho<1$

$$|\Phi(\zeta)|\leqq-\frac{D_p}{B_p'}\log(1-\rho). \tag{31.4}$$

The function $\Phi(\rho\zeta)$ is analytic in $|\zeta|<1$ and has there the bound indicated by (31.4). By §19, Corollary 3, we may write,

$$\rho|\Phi'(0)|+\frac{\rho^2}{2!}|\Phi''(0)|+\cdots+\frac{\rho^p}{p!}|\Phi^{(p)}(0)|$$

$$\leqq 24p\left[-\frac{D_p}{B_p'}\log(1-\rho)\right]^{1-2^{-p}}\cdot[D_p(0)]^{2^{-p}},$$

and this inequality is valid whether $D_p(0)$ refers to $\Phi(\rho\zeta)$ in $|\zeta|<1$, to $\Phi(\zeta)$

in $|\zeta|<\rho$, or to $\Phi(\zeta)$ in $|\zeta|<1$. The first part of Theorem 11 follows at once, where $D_p(w)$ refers now to $f(z)$, by §2, Lemma 2. The latter part of Theorem 11 follows from the inequality for $\rho=1/2$

$$\rho^{-p}[-\log(1-\rho)]^{1-2^{-p}} < 2^p.$$

An obvious consequence of Theorem 11 is that under the conditions of that theorem $D_p(w_n)\to 0$ implies

$$f^{(k)}(z_n)(1-|z_n|^2)^k \to 0, \qquad k = 1, 2, \cdots, p,$$

where $w_n=f(z_n)$, $|z_n|<1$; this conclusion is valid even if $w_n\to\infty$.

32. **Comments on condition** $|z_n|\to 1$. In the major part of the present paper, so far as it deals with $D_1(w)$, we are concerned with a function $f(z)$ analytic for $|z|<1$ and the two conditions

$$D_1(w_n) \to 0, \qquad w_n = f(z_n), \tag{32.1}$$

$$f'(z_n)(1-|z_n|^2) \to 0. \tag{32.2}$$

In the present section we propose to study the further condition

$$|z_n| \to 1 \tag{32.3}$$

in its relation to (32.1) and (32.2). To some extent, our remarks will be a recapitulation of material already developed.

The relation (32.2) implies (32.1) with no further restriction on $f(z)$, as follows from §4, Theorem 2.

For a univalent function $f(z)$, relation (32.1) implies (32.2) by §4, Theorem 1′. For such a function each of the conditions (32.1) and (32.2) implies (32.3), because $f'(z)$ has a positive lower bound in the closed region $|z|\leqq r<1$; but (32.3) does not imply (32.1) or (32.3), as is illustrated by the function $f(z)=z/(1-z)^2$, when real $z\to 1$; nevertheless(32.3) combined with the boundedness of $w_n$ implies (32.1) and (32.3), as follows by the kind of reasoning about to be given.

However, if $f(z)$ is both univalent and bounded, each of the conditions (32.1), (32.2), (32.3) implies all those conditions; it is sufficient now to show that (32.3) implies (32.1). The plane region $R$ which is the image of $|z|<1$ under the map $w=f(z)$ can be considered the sum of the plane regions $R_\nu$, the respective images of $|z|<1-1/\nu$, $\nu=1, 2, \cdots$, under the map $w=f(z)$. The regions $R_\nu$ increase monotonically; given an arbitrary $\delta>0$, there exists an index $N_\delta$ such that every point of $R_{N_\delta}$ lies within a distance less than $\delta$ of the boundary of $R$; the inequality $|z|>1-1/N_\delta$ implies $D_1(w)<\delta$; thus (32.3) implies (32.1) and hence (32.2).

Let now $f(z)$ be bounded in $|z|<1$; we have already indicated (§10) that the conditions (32.1) and (32.2) are equivalent. Nevertheless it is obvious that (32.1) does not imply (32.3); whenever $z_n$ approaches a point $z_0$ with $|z_0|<1, f'(z_0)=0$, the relation (32.1) is satisfied without (32.3); nevertheless,

if $D_1(w_n)\to 0$, there exists a subsequence of the $z_n$ which approaches a point $z_0$, with either $f'(z_0)=0$, $|z_0|<1$, or $|z_0|=1$. Reciprocally, Szegö's example (introduction to Chapter II) shows that (32.3) may be satisfied without (32.1).

Let us suppose now $f(z)$ bounded in $|z|<1$, $|f(z)|\leqq M$, $|z_n|\to 1$, $w_n=f(z_n)\to w_0$, $D_1(w_n)\geqq\delta>0$; we shall derive some geometric properties of the Riemann configuration $R$ onto which the transformation $w=f(z)$ maps $|z|<1$. By inequality (4.4) we may write also

$$|f'(z_n)|(1-|z_n|^2)\geqq\delta. \tag{32.4}$$

Let $r$ be arbitrary, $0<r<1$. The function

$$\phi(\zeta)=f\left(\frac{\zeta+z_n}{1+\bar{z}_n\zeta}\right)$$

is analytic in $|\zeta|<r$, has a modulus there not greater than $M$, with $|\phi'(0)|=|f'(z_n)|(1-|z_n|^2)\geqq\delta$. It follows from the Landau-Dieudonné theorem (§10) that the image of $|\zeta|<r$ under the transformation $w=f(z)$ contains a smooth circle whose center is $w_n$ and whose radius is at least $r^2\delta^2/8M=\delta_1$. By virtue of the relation $|z_n|\to 1$, it is possible to choose a subsequence $z_{n_k}$ having the property that the circle whose non-euclidean center is $z_{n_k}$ and non-euclidean radius 2 log $[(1+r)/(1-r)]$ contains on or within it none of the points $z_{n_{k+j}}$, $j>0$; as a consequence it follows from the triangle inequality that the circles $\gamma_{n_k}$ whose non-euclidean centers are the points $z_{n_k}$ having the common non-euclidean radius log $[(1+r)/(1-r)]$ are mutually exterior; this circle $\gamma_{n_k}$ is the image of $|\zeta|=r$ under the transformation $z=(\zeta+z_{n_k})/(1+\bar{z}_{n_k}\zeta)$. *Then the closed interiors of the smooth circles $C_{n_k}$ on $R$ whose centers are the respective points $w_{n_k}$ having the common radius $\delta_1$ are mutually disjoint.* By virtue of our assumption $w_n\to w_0$, it appears that the configuration $R$ has an infinity of separate sheets over the point $w=w_0$, each sheet containing a circle of center $w_0$ and radius $\delta_1-\eta$, where $\eta$ is arbitrary. We shall prove

THEOREM 12. *Let the function $f(z)$ analytic and bounded in $|z|<1$ admit a sequence $z_n$ with $|z_n|<1$, $|z_n|\to 1$,*

$$D_1(w_n)\geqq\delta>0, \qquad w_n=f(z_n). \tag{32.5}$$

*Then there exists a value $w=w_0$ such that the Riemann configuration $R$ onto which the transformation $w=f(z)$ maps $|z|<1$ has an infinity of separate sheets over the point $w=w_0$, each sheet containing a smooth circle whose center lies over the point $w=w_0$ and whose radius is $\delta_2>0$, where $\delta_2$ is suitably chosen.*

In Theorem 12, the condition (32.5) may of course be replaced by the condition that $|f'(z_n)|(1-|z_n|^2)$ should be bounded from zero, a condition that implies (32.5).

To prove Theorem 12 it suffices to apply the reasoning already given to a subsequence of the $w_n$ possessing a limit. Of course it is not possible to assert

here that the original sequence of circles of radii $D_1(w_n)$ corresponds to separate sheets of $R$; if the circles of radii $D_1(w_{2n})$ are given arbitrarily, corresponding to separate sheets of $R$, the point $z_{2n+1}$ can be chosen so near $z_{2n}$ that the corresponding circles overlap, while an inequality of form (32.5) persists.

Conversely, let $f(z)$ now be analytic and bounded for $|z|<1$, and let $w=f(z)$ map $|z|<1$ onto a Riemann configuration which has an infinity of separate sheets over some point $w_0$, each sheet containing a smooth circle $\gamma_n$ whose center lies over the point $w=w_0$ and whose radius is $\delta_2>0$; it is obvious that the centers of these smooth circles can be chosen as points $w_n$ so that the relation $D_1(w_n)\geqq\delta_2$ is fulfilled. The relation $|z_n|\to 1$ follows because otherwise a subsequence $z_{n_k}$ has a limit point $z_0$, with $|z_0|<1$; we have $w_0=f(z_{n_k})$, hence $w_0=f(z_0)$; an infinity of points $z_{n_k}$ lie in an arbitrary neighborhood of $z_0$; an infinity of the points $w_{n_k}=f(z_{n_k})$ on $R$ lie on $R$ in each $C_p$ whose center is $w_0=f(z_0)$, where $p-1$ is the order of $z_0$ as a zero of $f(z)$; this is in contradiction to our hypothesis that the $\gamma_n$ lie in distinct sheets of $R$; the converse of Theorem 12 is established.

In Theorem 12 and its converse, we have supposed $f(z)$ to be bounded; it also sufficient if $f(z)$ has two exceptional values in $|z|<1$; compare §22.

We add one further remark, in a somewhat different order of ideas. Let $w=f(z)$ be analytic in $|z|<1$, and let us suppose

$$\limsup_{z\to 1} D_1(w) < \infty; \tag{32.6}$$

this condition is a consequence of

$$\limsup_{z\to 1} |f'(z)|(1-|z|^2) < \infty, \tag{32.7}$$

if (32.7) itself is valid. It follows from (32.6) that $D_1(w)$ is uniformly bounded in $|z|<1$. Hence (32.1) and (32.2) are equivalent. Moreover, the discussion of Theorem 12 and its converse applies here. But even under these circumstances it is not true that (32.3) implies (32.1) or (32.2); this is shown by the function $w=f(z)$ with $f(0)=0$, $f'(0)>0$, which maps $|z|<1$ onto the strip $|v|<\pi$, where $w=u+iv$; we have $D_1(w)\leqq\pi$. But when $z_n$ is positive, $z_n\to 1$, we have $D_1(w_n)=\pi$, so neither (32.1) nor (32.2) is satisfied.

33. **$p$-valent functions.** For $p$-valent functions we can obtain results analogous to those for bounded functions and for functions which have two exceptional values.

THEOREM 13. *Let the function $f(z)$ be analytic and $p$-valent in the region $|z|<1$. Then we have*

$$|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)| \leqq A_p\cdot D_p(0), \tag{33.1}$$

*where $A_p$ is a numerical constant depending only on $p$.*

We assume $f(0)=0$, which obviously involves no loss of generality. We write for reference the inequality

$$\mu_p = \max [|a_1|, |a_2|, \cdots, |a_p|] \leqq |a_1| + |a_2| + \cdots + |a_p|,$$

$$a_k = \frac{1}{k!} f^{(k)}(0). \tag{33.2}$$

A theorem due to M. L. Cartwright([70]) asserts that under the conditions of Theorem 1, since we have $f(0)=0$, we have

$$|f(z)| \leqq A_p' \cdot \mu_p \cdot (1-r)^{-2p}, \qquad |z| \leqq r < 1, \tag{33.3}$$

where $A_p'$ is a number depending only on $p$ and where $\mu_p$ is defined by (33.2). We shall use (33.3) for the particular value $r=1/2$:

$$|f(z)| \leqq 2^{2p} \cdot A_p' \cdot \mu_p, \qquad |z| \leqq 1/2. \tag{33.4}$$

The function $F(z) \equiv f(z/2)$ is analytic in the region $|z| < 1$ and has there the bound $2^{2p} \cdot A_p' \cdot \mu_p$. If $D_p(0)$ refers to $F(z)$ or to $f(z)$ we have by §19, Corollary 3,

$$\begin{aligned} &|F'(0)| + \frac{1}{2!}|F''(0)| + \cdots + \frac{1}{p!}|F^{(p)}(0)| \\ &\qquad \leqq B_p[2^{2p} \cdot A_p' \cdot \mu_p]^{1-2^{-p}} \cdot [D_p(0)]^{2^{-p}} \end{aligned} \tag{33.5}$$

where $B_p$ may be chosen as $24p$. The first member of (33.5) can be written

$$\frac{1}{2}|f'(0)| + \frac{1}{2^2 \cdot 2!}|f''(0)| + \cdots + \frac{1}{2^p \cdot p!}|f^{(p)}(0)|,$$

which is not greater than

$$\frac{1}{2^p}\left[|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)|\right].$$

A consequence of (33.5) and (33.2) is then the inequality

$$\begin{aligned} &\left[|f'(0)| + \frac{1}{2!}|f''(0)| + \cdots + \frac{1}{p!}|f^{(p)}(0)|\right]^{2^{-p}} \\ &\qquad \leqq 2^p \cdot B_p[2^{2p} \cdot A_p']^{1-2^{-p}} \cdot [D_p(0)]^{2^{-p}}, \end{aligned}$$

which can be put into the form (33.1).

By virtue of §2, Lemma 2 and §20, Theorem 2, we can formulate from Theorem 1

THEOREM 14. *Let the function $f(z)$ be analytic and $p$-valent in the region $|z|<1$. Then with the conditions $|z_0|<1$, $w_0=f(z_0)$, we have*

([70]) Mathematische Annalen, vol. 111 (1935), pp. 98–118.

$$\gamma_p \cdot D_p(w_0) \leqq \sum_{k=1}^{p} \left| \sum_{\nu=0}^{k-1} (-1)^\nu C_{k-1,\nu} \bar{z}_0^\nu \frac{(1-|z_0|^2)^{k-\nu}}{(k-\nu)!} \cdot f^{(k-\nu)}(z_0) \right| \leqq \Theta_p \cdot D_p(w_0),$$

*where* $\gamma_p$ *is the number of* §20, *Theorem* 2, *and* $\Theta_p$ *is a number depending only on* $p$ *which may be chosen as* $A_p$ *in Theorem* 1.

*Consequently if we have* $|z_n| < 1$, $w_n = f(z_n)$, *a necessary and sufficient condition for*

$$\lim_{n\to\infty} f^{(k)}(z_n)(1 - |z_n|^2)^k = 0, \qquad k = 1, 2, \cdots, p,$$

*is the condition* $\lim_{n\to\infty} D_p(w_n) = 0$.

The case $p=1$ brings us back to §4, Theorem 3.

34. **Some extensions to meromorphic functions.** Let us consider a class of functions $f(z)$ meromorphic in $|z|<1$, omitting there the three distinct values $a, b, c$, and such that $f(0) = A$, $|A| \leqq A_0$, where $A_0$ is a positive constant independent of the particular function of the class. Corresponding to this class there exists a number $\theta$ $(0<\theta<1)$ such that we have

$$|f(z)| \leqq \Omega(A_0, \theta) \tag{34.1}$$

for $|z|<\theta$, where $\Omega$ is independent of any particular function of the class.

Indeed, suppose no such value of $\theta$ existed. On the circle $|z| = 1/n$ some function $f_n(z)$ would attain a value of modulus exceeding $n$. From the sequence of functions $f_n(z)$ one can extract a subsequence converging uniformly[71] in every closed subregion of $|z|<1$ either to a meromorphic function or to the infinite constant. The second alternative cannot take place since by hypothesis $|f_n(0)| \leqq A_0$ for all $n$. But, on the other hand, if the sequence $f_n(z)$ converges to a meromorphic function, the latter must have a pole at the origin which is not possible on account of the condition $|f_n(0)| \leqq A_0$. Hence, the asserted existence of $\theta$ has been established.

Let $z_0$ $(|z_0|<1)$ be a point such that, setting $w_0 = f(z_0)$, we have $|w_0| \leqq A_0$. Consider the function

$$\phi(\zeta) = f\left(\frac{\zeta + z_0}{1 + \bar{z}_0\zeta}\right)$$

which is meromorphic in $|\zeta|<1$, omits there the values $a, b, c$, and for which $\phi(0) = w_0$. In accordance with (34.1) we have

$$|\phi(\zeta)| \leqq \Omega(A_0, \theta)$$

in $|\zeta|<\theta$. Hence, in $|\zeta|<1$ we have

$$|\phi(\theta\zeta)| \leqq \Omega(A_0, \theta).$$

---

(71) Defined, for instance, as by Montel, *Leçons sur les Familles Normales de Fonctions Analytiques*, Paris, 1927, p. 124.

Now, applying Theorem 5, Chapter III, we obtain the inequality

$$(34.2)\qquad \theta\,|\,\phi'(0)\,| + \frac{\theta^2}{2!}\,|\,\phi''(0)\,| + \cdots + \frac{\theta^p}{p!}\,|\,\phi^{(p)}(0)\,| \leqq \Lambda_p'\,[D_p(w_0)]^{2-p},$$

where $\Lambda_p'$ depends on $p$, $\theta$, and $A_0$. It is clear, furthermore, that $\theta$ depends on $a$, $b$, $c$, $A_0$ but not on $\phi(\zeta)$ and consequently may be omitted by modifying $\Lambda_p'$ properly. It is also to be noted that $D_p(w_0)$ in (34.2) is the radius of $p$-valence at the point $w_0$ of the Riemann surface on which the function $\phi(\theta\zeta)$ maps $|\zeta|<1$ which is the same as the radius of $p$-valence at the point $w_0$ of the Riemann surface on which $\phi(\zeta)$ maps $|\zeta|<\theta$. This radius of $p$-valence is not greater than the radius of $p$-valence at the point $w_0$ of the Riemann surface on which $\phi(\zeta)$ maps $|\zeta|<1$. Hence, if in (34.2) we return to the function $f(z)$ we obtain the inequality

$$(34.3)\qquad \sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^{\nu}C_{k-1,\nu}\bar{z}_0^{\nu}\,\frac{(1-|\,z_0\,|^2)^{k-\nu}f^{(k-\nu)}(z_0)}{(k-\nu)!}\right| \leqq \Lambda_p'\,[D_p(w_0)]^{2-p},$$

where $D_p(w_0)$ is now the radius of $p$-valence at the point $w_0$ of the Riemann surface on which $f(z)$ maps the circle $|z|<1$.

Now, as is remarked in §21 after the proof of Theorem 2, that theorem requires analyticity only in the neighborhood of the origin, which $f(z)$ possesses in $|z|<\theta$. Hence, applying Theorem 2 we find that

$$(34.4)\qquad \lambda_p D_p(w_0) \leqq \sum_{k=1}^{p}\left|\sum_{\nu=0}^{k-1}(-1)^{\nu}C_{k-1,\nu}\bar{z}_0^{\nu}\,\frac{(1-|\,z_0\,|^2)^{k-\nu}f^{(k-\nu)}(z_0)}{(k-\nu)!}\right|,$$

where $\lambda_p$ depends on $p$ alone. Thus, we may state

THEOREM 15. *Let $f(z)$ be a function meromorphic in $|z|<1$, omitting there the three distinct values $a$, $b$, $c$. Let $z_0$ ($|z_0|<1$) be a point such that, setting $w_0=f(z_0)$, $|w_0|\leqq A_0$ where $A_0$ is a positive constant. Then, the inequalities* (34.3) *and* (34.4) *hold, where $\Lambda_p'$ depends on $A_0$, $a$, $b$, $c$, but not on $z_0$ or $f(z)$.*

It follows that if under the hypotheses of Theorem 15 for a sequence of points $z_n$ ($|z_n|<1$) the sequence $w_n=f(z_n)$ is bounded, then a necessary and sufficient condition for $\lim_{n\to\infty} f^{(k)}(z_n)\,(1-|z_n|^2)^k=0$ ($k=1, 2, \cdots, p$) is $\lim_{n\to\infty} D_p(w_n)=0$.

It will be noted that under the conditions of Theorem 15 we have $D_p(w)\leqq|w-a|$ so that inequality (34.3) gives an inequality on the approach to zero of $(1-|z|^2)^k|f^{(k)}(z)|$ as $w$ tends to zero, for every $k$.

We add the remark that much of the discussion of §27 can be carried over to meromorphic functions which omit three values; this development is left to the reader.

UNIVERSITY OF ROCHESTER,
ROCHESTER, N. Y.
HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

# ANALYTIC EXTENSION BY HAUSDORFF METHODS

BY

RALPH PALMER AGNEW

**1. Introduction.** Let $\chi(t)$ be a complex-valued function having bounded variation over $0 \leqq t \leqq 1$, let $\chi(0)=0$, and let $\chi(t)$ have no removable discontinuities. This *mass function* $\chi(t)$ generates a *Hausdorff transformation* (Hausdorff [9]) $H(\chi)$

$$\sigma_n = \int_0^1 \sum_{k=0}^{n} C_{n,k} t^k (1-t)^{n-k} s_k \, d\chi(t)$$

by means of which a series $u_0+u_1+u_2+\cdots$ with partial sums $s_0, s_1, s_2, \cdots$ is summable $H(\chi)$ to $\sigma$ if $\sigma_n \to \sigma$ as $n \to \infty$. The transformation $H(\chi)$ is regular (that is, such that existence of lim $s_n$ implies lim $\sigma_n$ = lim $s_n$) if and only if $\chi(1)=1$ and $\chi(t)\to 0$ as $t \to 0$.

Among the familiar regular transformations obtained by specializing $\chi(t)$ are the methods $C_r$ ($\Re(r)>0$) of Cesàro for which $\chi(t)=1-(1-t)^r$; the methods $H_r$ ($\Re(r)>0$) of Hölder for which

$$\chi(t) = \frac{1}{(r-1)!}\int_0^t \left(\log \frac{1}{u}\right)^{r-1} du;$$

and the methods $E_r$ ($0<r\leqq 1$) of Euler for which $\chi(t)=0$ or 1 according as $0 \leqq t < r$ or $r \leqq t \leqq 1$.

It is well known that neither Cesàro nor Hölder methods are effective in evaluating power series outside their circles of convergence. Characterization of the region of the complex plane in which a power series $\sum a_n z^n$ is summable $E_r$ was given by Knopp [14] and Rademacher [18] for the cases in which $r$ is of the form $2^{-k}$, $k=1, 2, 3, \cdots$. If $z_0$ is an interior point of the Borel polygon determined by $\sum a_n z^n$, then $\sum a_n z_0^n$ is summable $E_r$ provided $r$ lies within a sufficiently small interval $0<r<\delta$ to the right of the origin. On the other hand, if $z_1$ is a point outside the Borel polygon, then no regular Euler method can evaluate $\sum a_n z_1^n$. Beyond these facts and a few corollaries of them, very little seems to be known about the problem of analytic extension by means of Hausdorff methods of summability. It has been conjectured by Garabedian and Wall [8] that a regular Hausdorff method is ineffective outside the circle of convergence unless $\chi(t)$ is constant over some interval $1-\delta \leqq t \leqq 1$. For references to literature (up to 1927) on analytic extension by various methods, see Hille [10].

Presented to the Society, September 5, 1941; received by the editors July 14, 1941.

Long ago it became the fashion to test the efficacy of methods of summability by applying them to the geometric series $\sum z^n$. We determine, in terms of the generating function $\chi(t)$, the open region of the $z$-plane in which $\sum z^n$ is summable by a regular Hausdorff method $H(\chi)$. The result is set forth in the following theorem.

THEOREM 1.1. *Let $\chi(t)$ have bounded variation over $0 \leqq t \leqq 1$, let $\chi(0+) = \chi(0) = 0$, let $\chi(1) = 1$, and let $\chi(t)$ have no removable discontinuities. Let $r$ be the greatest lower bound of all numbers $\rho$ such that $\chi(t) = 1$ when $\rho \leqq t \leqq 1$. Then $\sum z^n$ is summable $H(\chi)$ to $1/(1-z)$ at each point $z$ interior to the circle*

$$\left| z - \left(1 - \frac{1}{r}\right) \right| = \frac{1}{r}$$

*and $\sum z^n$ is non-summable $H(\chi)$ at each point $z$ exterior to the same circle.*

Let the number $r$ described in the statement of the theorem be termed the *order* of the transformation $H(\chi)$, and let the circle be termed the *circle of summability*. If (as is the case for the Cesàro and Hölder methods) there is no interval $\rho < t \leqq 1$ over which $\chi(t)$ is constant, then the order $r$ of $H(\chi)$ is 1 and the circle of summability is identical with the circle of convergence. If (as is the case for the Euler methods of order $0 < r < 1$) there is an interval $\rho < t \leqq 1$ over which $\chi(t)$ is constant, then the order $r$ of $H(\chi)$ is less than 1. The order must be positive since $\chi(t) \to 0$ as $t \to 0$. Thus the radius $r^{-1}$ of the of the circle of summability is greater than 1, the circle of summability being tangent to the line $\Re(z) = 1$ at the point $z = 1$. Since the center and radius of the circle of summability depend only upon the order of $H(\chi)$, it follows that *two Hausdorff methods of equal order have the same circle of summability*. In particular, the circle of summability of a Hausdorff method of order $r$ is the same as that of the Euler method $E_r$ of equal order(1). Differences in effectiveness of two Hausdorff methods of equal order can appear only for points $z$ on the circle of summability. It is a consequence of Theorem 1.1 that the effectiveness of methods $H(\chi)$ in evaluating $\sum z^n$ increases steadily as the order $r$ decreases. Each fixed point $z_0$ for which $\Re z_0 < 1$ lies within the circle of summability of $H(\chi)$ provided the order $r$ of $H(\chi)$ is sufficiently near 0. Each fixed point $z_1$ for which $z_1 \neq 1$ and $\Re z_1 \geqq 1$ lies outside each circle of summability and accordingly $\sum z_1^n$ is non-summable $H(\chi)$ for each regular $H(\chi)$.

Proof of summability of $\sum z^n$ when $z$ lies inside the circle of summability is very simple and is given in §2. Proof of non-summability is more complicated; this and related facts are proved in §3. In §§4 and 5, we establish uniform summability of power series over appropriate sets. In §§6 and 7, we

(1) A regular method $H(\chi)$ includes $E_r$ if and only if the order of $H(\chi)$ is less than or equal to $r$. See Hille and Tamarkin [11]. For a general discussion of inclusion relations involving regular Hausdorff methods, see Garabedian, Hille, and Wall [7].

define and discuss *collective Hausdorff summability*. In §8, we improve a known theorem relating Abel and Hausdorff summability.

**2. Proof of summability.** When $z\neq 1$, the series $\sum z^n$ has partial sums

$$s_k = (1 - z^{k+1})/(1 - z), \qquad k = 0, 1, 2, \cdots,$$

and the $H(\chi)$ transform of the series $\sum z^n$ is accordingly

$$\begin{aligned}\sigma_n(z) &= \frac{1}{1-z} - \frac{z}{1-z}\int_0^1 \sum_{k=0}^{n} C_{n,k}(tz)^k(1-t)^{n-k}d\chi(t) \\ &= \frac{1}{1-z} - \frac{z}{1-z} f_n(z)\end{aligned} \tag{2.1}$$

where

$$f_n(z) = \int_0^1 [1 + t(z-1)]^n d\chi(t). \tag{2.2}$$

The sequence $f_n(z)$ is the $H(\chi)$ transform of the sequence $s_k \equiv z^k$.

Let $\chi(t)$ be regular and let $z_0$ be a fixed point inside the circle of summability; we show that $\sum z_0^n$ is summable to $1/(1-z_0)$ by showing that $f_n(z_0)\rightarrow 0$ as $n\rightarrow\infty$. Since $z_0$ lies inside the circle of summability,

$$|1 + r(z_0 - 1)| < 1 \tag{2.3}$$

where $r$ is the order of $H(\chi)$ and accordingly $\chi(t) = 1$ when $r < t \leqq 1$. We may assume that $\chi(t) = 1$ over $r \leqq t \leqq 1$. Let $\epsilon > 0$. Then

$$\begin{aligned}|f_n(z_0)| &= \left|\int_0^r [1 + t(z_0 - 1)]^n d\chi(t)\right| \\ &\leqq \int_0^\delta |1 + t(z_0-1)|^n |d\chi(t)| + \int_\delta^r |1 + t(z_0-1)|^n |d\chi(t)| \\ &\leqq \int_0^\delta |d\chi(t)| + \int_\delta^r A^n |d\chi(t)| < \epsilon\end{aligned} \tag{2.4}$$

when $n$ is sufficiently great, provided $\delta$ is a positive number so chosen that

$$\int_0^\delta |d\chi(t)| < \epsilon/2, \tag{2.5}$$

and $A$ is a constant for which

$$\max_{\delta \leqq t \leqq r} |1 + t(z_0 - 1)| \leqq A < 1. \tag{2.6}$$

Hence $f_n(z_0)\rightarrow 0$ and our result follows. In case $F$ is a closed set interior to the circle of summability, it is possible to fix $\delta > 0$ such that (2.5) holds and

then show existence of a constant $A$ such that (2.6) holds for each $z_0 \in F$; this implies uniform summability over the set $F$.

3. **Proof of non-summability.** Assuming that $\chi(t)$ satisfies the hypotheses of Theorem 1.1 and that $z$ is a fixed point outside the circle of summability, we show that $\sum z^n$ is not summable $H(\chi)$. For this, it is sufficient to show that the $H(\chi)$ transform $f_n(z)$ of the sequence $z^n$ is unbounded.

The hypothesis that $z$ lies outside the circle of summability implies that $|z| > 1$. Hence there is a number $t_1$ such that $0 \leqq t_1 < 1$ and

$$\begin{aligned} |1 + (z-1)t| &\leqq 1, \qquad & 0 \leqq t \leqq t_1, \\ &> 1, & t_1 < t \leqq 1. \end{aligned}$$

In case $\Re z \geqq 1$, $t_1$ is 0; and in case $\Re z < 1$, we have $0 < t_1 < 1$. Again using the hypothesis that $z$ lies outside the circle of summability, we obtain

$$|1 + (z-1)r| > 1$$

where $r$ is the order of $H(\chi)$, and accordingly $r > t_1$. Hence our result is a consequence of the following theorem.

THEOREM 3.1. *If $|z| > 1$ and $\chi(t)$ is a function of bounded variation over $0 \leqq t \leqq 1$ having no removable discontinuities, then a necessary and sufficient condition that the $H(\chi)$ transform*

$$f_n(z) = \int_0^1 [1 + (z-1)t]^n d\chi(t)$$

*of the sequence $s_k \equiv z^k$ be bounded is that $\chi(t)$ be constant over the interval $t_1 < t \leqq 1$ over which $|1+(z-1)t| > 1$*(2).

Sufficiency is obvious; for if $\chi(t)$ is constant over $t_1 < t \leqq 1$, then we may assume $\chi(t)$ constant over $t_1 \leqq t \leqq 1$ and obtain

$$\begin{aligned} |f_n(z)| &\leqq \int_0^1 |1 + (z-1)t|^n \, |d\chi(t)| \\ &= \int_0^{t_1} |1 + (z-1)t|^n \, |d\chi(t)| \leqq \int_0^{t_1} |d\chi(t)|. \end{aligned}$$

To prove necessity, we simplify writing by setting $\zeta = z - 1$, and assume that the sequence $f_n$ defined by

$$f_n = \int_0^1 (1 + \zeta t)^n d\chi(t)$$

(2) In case $H(\chi)$ is an Euler method $E_r$ for which $0 < r < 1$, $f_n(z) = [1 + (z-1)r]^n$ and we see immediately that $f_n(z)$ is bounded if and only if $|1+(z-1)r| \leqq 1$ and hence if and only if $0 < r \leqq t_1$ and therefore if and only if $\chi(t) = 1$ over $t_0 < t \leqq 1$. The case in which $H(\chi)$ is not an Euler method is not so simple.

is bounded. Integrating by parts, setting

$$u(t) = (1 + \zeta t)^n, \qquad v_1(t) = -\int_t^1 d\chi(s) = \chi(t) - \chi(1),$$

we obtain

$$f_n = \int_0^1 d\chi(t) - n\zeta \int_0^1 (1 + \zeta t)^{n-1} v_1(t) dt. \tag{3.11}$$

If we show that $v_1(t) = 0$ over $t_1 < t \leqq 1$, our result will follow. The function $v_1(t)$ has bounded variation over $0 \leqq t \leqq 1$ and has no removable discontinuities. If we set

$$I_n^{(1)} = \int_0^1 (1 + \zeta t)^n v_1(t) dt, \tag{3.12}$$

then (3.11) and our hypotheses imply that $I_n^{(1)} \to 0$ as $n \to \infty$. Integrating by parts, setting

$$u = (1 + \zeta t)^n, \qquad v_2(t) = -\int_t^1 v_1(u) du,$$

we obtain

$$I_n^{(1)} = \int_0^1 v_1(t) dt - n\zeta \int_0^1 (1 + \zeta t)^{n-1} v_2(t) dt.$$

If we show that $v_2(t) = 0$ over $t_1 < t \leqq 1$, our result will follow. The function $v_2(t)$ is continuous and if we set

$$I_n^{(2)} = \int_0^1 (1 + \zeta t)^n v_2(t) dt$$

then $I_n^{(2)} \to 0$ as $n \to \infty$. If we integrate by parts once more, setting

$$g(t) = -\int_t^1 v_2(u) du,$$

we see that our result is a consequence of the following lemma.

LEMMA 3.2. *If $g(t)$ has a continuous derivative, if $|1 + \zeta| > 1$, and if*

$$I_n = \int_0^1 (1 + \zeta t)^n g(t) dt, \qquad n = 0, 1, 2, \cdots, \tag{3.21}$$

*is bounded, then $g(t) = 0$ over the interval $t_1 < t \leqq 1$ of values of $t$ for which $|1 + \zeta t| > 1$.*

Since $I_n$ is bounded, the series $\sum I_n w^n$ converges and defines a function

$$F_1(w) = \sum_{n=0}^{\infty} I_n w^n$$

analytic over the open circular region $|w| < 1$ of a complex $w$-plane. If $|w| < 1/|1+\zeta|$, then

$$|w(1 + \zeta t)| \leqq |w||1 + \zeta| < 1, \qquad 0 \leqq t \leqq 1,$$

so that

$$\sum_{n=0}^{\infty} (1 + \zeta t)^n w^n g(t)$$

converges uniformly over $0 \leqq t \leqq 1$. Hence it follows from (3.21) that when $|w| < 1/|1+\zeta|$,

$$\sum_{n=0}^{\infty} I_n w^n = \int_0^1 \frac{g(t)}{1 - w - w\zeta t}\, dt.$$

But the function $F_2(w)$ defined by

$$F_2(w) = \int_0^1 \frac{g(t)}{1 - w - w\zeta t}\, dt$$

is analytic at each point $w$ of the complex plane except possibly those for which the equation

$$w = \frac{1}{1 + \zeta t}$$

holds for some $t$ in the interval $0 \leqq t \leqq 1$.

Let $t_0$ be fixed such that $0 < t_0 \leqq 1$ and $|1+\zeta t_0| > 1$. If we set, for $y$ real,

$$w_y = \frac{1}{1 + \zeta(t_0 + iy)}$$

then

$$1 - w_y(1 + \zeta t) = 1 - \frac{1 + \zeta t}{1 + \zeta(t_0 + iy)} = \frac{\zeta(t_0 - t + iy)}{1 + \zeta(t_0 + iy)}$$

and hence when $y \neq 0$

$$F_2(w_y) = -\frac{1 + \zeta(t_0 + iy)}{\zeta} \int_0^1 \frac{g(t)}{t - t_0 - iy}\, dt.$$

But $F_1(w)$ and $F_2(w)$ are functions which are analytic and equal at points $w = w_y$ for which $y$ is real, not 0, and $|y|$ is sufficiently small. Moreover $F_1(w)$ is analytic at the point $w = w_0 \equiv 1/(1+\zeta t_0)$. Therefore $\lim_{y \to 0} F_2(w_y)$ must exist. Hence if we set

$$G(y) = \int_0^1 \frac{g(t)}{t - t_0 - iy}\, dt, \qquad y \neq 0,$$

then $\lim_{y\to 0} G(y)$ must exist. We are now in a position to complete the proof by showing that $g(t_0)$ must be 0.

Since $g(t)$ has a continuous derivative, there is a continuous function $B(t)$ such that

$$g(t) = B(t)(t - t_0) + g(t_0).$$

Hence $G(y) = G_1(y) + G_2(y)$ where

$$G_1(y) = \int_0^1 \frac{B(t)(t - t_0)}{t - t_0 - iy}\, dt, \qquad G_2(y) = \int_0^1 \frac{g(t_0)}{t - t_0 - iy}\, dt.$$

Since the first integrand is dominated by $|B(t)|$ and converges to $B(t)$ as $y \to 0$, $\lim_{y\to 0} G_1(y)$ exists; hence $\lim_{y\to 0} G_2(y)$ must exist. Upon evaluating the integral for $G_2(y)$, we see that existence of $\lim_{y\to 0} G_2(y)$ implies that $g(t_0) = 0$. This completes the proof of Lemma 3.2 and hence also the proofs of Theorems 3.1 and 1.1.

4. **Uniform summability of power series inside Euler polygons $B(r)$.** Let $\sum c_n z^n$ be a power series with a finite positive radius of convergence $R$, and let $f(z)$ be the function generated by analytic extension along radial lines from the origin. The open set in which $f(z)$ is thus defined is the *Mittag Leffler star* $S$. If a half-line $l_\phi$ of points $z$, such that $z = \rho e^{i\phi}$ where $\rho \geqq 0$, contains no singular point of $f(z)$, then $l_\phi$ is in $S$. If $l_\phi$ contains a singular point, and $\rho_0$ is the least value of $\rho$ for which $\rho e^{i\phi}$ is a singular point, then the point $\zeta = \rho_0 e^{i\phi}$ is a *vertex* of the star; the points of $l_\phi$ for which $0 \leqq \rho < \rho_0$ lie in $S$ and the points for which $\rho \geqq \rho_0$ are exterior to $S$.

Let $r$ be fixed such that $0 < r \leqq 1$. Corresponding to each vertex $\zeta$ of the star, let $B(r, \zeta)$ denote the set of points $z$ for which

$$\left| z - \left(1 - \frac{1}{r}\right)\zeta \right| < \frac{|\zeta|}{r}. \tag{4.01}$$

This set $B(r, \zeta)$ is the interior of the circle, with center at $(1 - r^{-1})\zeta$, which passes through the point $\zeta$. The set $B(r, \zeta)$ contains the interior of the circle of convergence. Let $B(r)$ denote the set of inner points of the intersection of the sets $B(r, \zeta)$ determined by the set of vertices $\zeta$ of $S$. This set $B(r)$, which is not a polygon in the ordinary sense, was called a *curvilinear polygon* by Knopp [14]; we shall call it the *Euler polygon of order r*. For each $r$, $B(r)$ is a bounded convex open set containing the inner points of the circle of convergence. If $r_1 < r_2$, then $B(r_2, \zeta) \subset B(r_1, \zeta)$ for each $\zeta$ and accordingly $B(r_2) \subset B(r_1)$. The union, for $0 < r \leqq 1$ of the sets $B(r)$ is (Knopp [14] and Rademacher [18]) the Borel polygon $B$. It was shown by Knopp and Rademacher, for the case

in which $r=2^{-p}$, that $\sum c_n z^n$ is summable $E_r$ or non-summable $E_r$ according as $z$ lies inside or outside the Euler polygon $B_r$ of order $r$. The following theorem presents a fact, involving summability by regular Hausdorff methods, which is the analogue of the fundamental fact that a power series having a finite radius of convergence converges uniformly over each closed set $F$ inside the circle of convergence.

THEOREM 4.1. *If $H(\chi)$ is a regular Hausdorff transformation of order $r$ and $F$ is a closed subset of the Euler polygon $B(r)$ of order $r$ of a power series $\sum c_n z^n$ having a finite positive radius of convergence, then $\sum c_n z^n$ is summable $H(\chi)$ uniformly over $F$ to the function $f(z)$ obtained by analytic extension of $\sum c_n z^n$ along radial lines from the origin.*

Let $z_1 \in B(r)$. Then, when $\zeta$ is a vertex of the star $S$, $|z_1-(1-r^{-1})\zeta| < r^{-1}\zeta$. Let $\zeta'$ be a point not in $S$. Then a vertex $\zeta$ of $S$ and a number $\rho \geqq 1$ exist such that $\zeta'=\rho\zeta$. The circular set of points $z$ for which $|z-(1-r^{-1})\zeta'| < r^{-1}|\zeta'|$ contains the circular set of points $z$ for which $|z-(1-r^{-1})\zeta| < r^{-1}|\zeta|$ and hence contains $z_1$. This shows that if $z_1 \in B(r)$ and $\zeta'$ is not in the star, then $|z_1-(1-r^{-1})\zeta'| < r^{-1}|\zeta'|$. It follows that if $z_1 \in B(r)$, then the set of all points $u$ for which

$$|z_1 - (1 - r^{-1})u| \geqq r^{-1}|u| \tag{4.11}$$

must lie in the star. The set of points $u$ for which (4.11) holds is the set for which

$$\left| u - \frac{1-r}{2-r} z_1 \right| \leqq \frac{|z_1|}{2-r}. \tag{4.12}$$

Therefore, if $z_1 \in B(r)$, the circular set of points $u$ satisfying (4.11) and (4.12) must lie in the star $S$. This circular set contains the origin in its interior, and the point $z_1$ lies on the boundary.

Let $\rho(\phi)$ be the continuous positive function such that the point $w=\rho(\phi)e^{i\phi}$ traverses the boundary of $B(r)$ as $\phi$ increases from 0 to $2\pi$. Corresponding to each multiplier $h$ for which $0<h<1$, let $\beta(r, h)$ be the open set of points inside the curve $\omega = h\rho(\phi)e^{i\phi}$. The sets $\beta(r, h)$ are *nested* subsets of $B(r)$ in the sense that if $0<h_1<h_2<1$, then $\bar\beta(r, h_1) \subset \beta(r, h_2) \subset B(r)$. Since $F$ is a closed subset of the open set $B(r)$, there is a multiplier $h_0$ such that $0<h_0<1$ and $F \subset \beta(r, h_0)$. Let $h_1$ be fixed such that $h_0<h_1<1$. Since $H(\chi)$ is regular, the $H(\chi)$ transform of $\sum c_n z^n$ converges to $f(0)$ when $z=0$. Accordingly, it is sufficient to prove that the $H(\chi)$ transform of $\sum c_n z^n$ converges to $f(z)$ uniformly over the set $\beta_0$ obtained by deleting the point $z=0$ from the set $\beta(r, h_0)$.

Let $z \in \beta_0$. Then there is a number $\lambda \equiv \lambda(z) > 1$ such that the point $\lambda z$ lies on the boundary of the set $\beta(r, h_1)$; we shall want to use the fact that $\lambda > \lambda_0$ where $\lambda_0 = h_1/h_0$. Let $C(z)$ denote the circle composed of the points $u$ for which

$$\left| \lambda z - (1 - r^{-1})u \right| = r^{-1} \left| u \right|. \tag{4.21}$$

Since $\lambda z \in B(r)$, the circle $C(z)$ lies in the star. The equation of $C(z)$ can be written in the form

$$\left| u - \frac{1 - r}{2 - r} \lambda z \right| = \frac{\left| \lambda z \right|}{2 - r}. \tag{4.22}$$

From this equation, it is apparent that the origin and the point $z$ lie inside the circle $C(z)$; in fact 0 and $z$ are points of the diameter having its ends at the points $-[r/(2-r)]\lambda z$ and $\lambda z$. Since $0 < r/(2-r) < 1 < \lambda$, it is apparent that if $u \in C(z)$, then

$$\left| u \right| \geqq \frac{r}{2 - r} \left| \lambda z \right| \geqq \frac{rc_1}{2 - r} = c_2 > 0 \tag{4.23}$$

where $c_1$ is the minimum value of $\left| \lambda z \right|$ when $\lambda z$ lies on the boundary of the set $\beta(r, h_1)$ and accordingly $c_1 = h_1 R$ where $R$ is the radius of convergence of $\sum c_n z^n$; and where $c_2$ is defined by the last equality. Since on the one hand $\left| \lambda z - z \right| \geqq c_3 > 0$ where $c_3$ is the minimum value of $\left| z_1 - z_2 \right|$ when $z_1$ lies on the boundary of the set $\beta(r, h_1)$ and $z_2$ lies in the closure of the set $\beta(r, h_0)$, and on the other hand

$$\left| z - [-r/(2 - r)]\lambda z \right| \geqq \left| [r/(2 - r)]\lambda z \right| \geqq c_2,$$

it follows that when $u \in C(z)$

$$\left| u - z \right| \geqq c_4 > 0, \tag{4.24}$$

the constant $c_4$ being the minimum value of $c_3$ and $c_2$.

When $z \in \beta_0$ and $u \in C(z)$, the equality (4.21) gives

$$\left| \frac{z}{u} - \frac{1}{\lambda}\left(1 - \frac{1}{r}\right) \right| = \frac{1}{\lambda} \cdot \frac{1}{r}. \tag{4.31}$$

If $\mu_1 \leqq \mu_0$, then the set of points $v$ of the complex plane for which the inequality

$$\left| v - \mu(1 - r^{-1}) \right| \leqq \mu r^{-1} \tag{4.32}$$

holds when $\mu = \mu_1$ is a subset of the set of points $v$ for which the inequality holds when $\mu = \mu_0$. Hence we can use (4.31) and the fact that $\lambda \equiv \lambda(z) \geqq \lambda_0$ to obtain

$$\left| \frac{z}{u} - \frac{1}{\lambda_0}\left(1 - \frac{1}{r}\right) \right| \leqq \frac{1}{\lambda_0} \cdot \frac{1}{r}.$$

When $\mu$ has the fixed value $\mu = \lambda_0^{-1} < 1$, the set of points $v$ for which (4.32) holds is a closed subset of the set of points $v$ for which

$$\left| v - (1 - r^{-1}) \right| < r^{-1}.$$

Hence there is a constant $\theta$, depending only on $r$ and $\lambda_0$ and independent of $z$ and $u$, such that $0<\theta<1$ and

$$\left|\frac{z}{u}-\left(1-\frac{1}{r}\right)\right| \leqq \frac{\theta}{r}\cdot \tag{4.33}$$

If $\delta$ is a fixed number for which $0<\delta<r$, then (4.33) implies existence of $\alpha$ and $\phi$ such that $0<\alpha\leqq 1$, $0<\phi\leqq 2\pi$ and

$$\frac{z}{u}=\left(1-\frac{1}{r}\right)+\frac{\alpha\theta}{r}e^{i\phi}$$

and it follows easily that, when $\delta\leqq t\leqq r$,

$$\begin{aligned}\left|1+t\left(\frac{z}{u}-1\right)\right| &= \left|1-\frac{t}{r}+\frac{\alpha t\theta}{r}e^{i\phi}\right| \leqq 1-t\left(\frac{1-\theta}{r}\right)\\ &\leqq 1-\delta\left(\frac{1-\theta}{r}\right)=\theta_1<1\end{aligned} \tag{4.34}$$

where $\theta_1$ is a constant, defined by the equality, depending on $\delta$, $\theta$, and $r$ but independent of $z$ and $u$. These considerations show also that

$$\left|1+t\left(\frac{z}{u}-1\right)\right| \leqq 1, \qquad 0\leqq t\leqq r. \tag{4.35}$$

For each $h$ in the interval $0<h<1$, let $W(h)$ denote the set which contains those and only those points $w$ for which the inequality

$$\left|z'-(1-r^{-1})w\right| \geqq r^{-1}\left|w\right| \tag{4.41}$$

holds for at least one point $z'$ in the closure of the set $\beta(r, h)$ defined above. For each $h$, $W(h)$ is [see (4.11)] the union of sets in the star and accordingly $W(h)$ is in the star. It is probably true that $W(h)$ is closed, but we do not need the result. For each $h$, the set $W(h)$ is bounded. If $0<h_1<h_2<1$, then the points of the closure of $W(h_1)$ are inner points of $W(h_2)$ and it follows that the closure $\overline{W}(h_1)$ of $W(h_1)$ lies in the star. Since $\overline{W}(h_1)$ is a bounded closed subset of the star, $f(z)$ must be bounded over $\overline{W}(h_1)$. Choose a constant $c_5$ such that

$$\left|f(z)\right| \leqq c_5, \qquad z\in\overline{W}(h_1).$$

From the definitions of $C(z)$ and $W(h_1)$, it follows that the points on the curve $C(z)$ lie in the set $W(h_1)$ for each $z\in\beta_0$. Hence, when $z\in\beta_0$,

$$\left|f(u)\right| \leqq c_5, \qquad u\in C(z), \tag{4.42}$$

the constant $c_5$ being independent of $z$ and $u$.

For each $z \in \beta_0$, let $C(z)$ be the circle defined above. The coefficients $c_j$ are related to $f(z)$ by the familiar formula

$$c_j = \frac{1}{2\pi i}\int_{C(z)} \frac{f(u)}{u^{j+1}}\,du, \qquad j = 0, 1, 2, \cdots; \tag{4.51}$$

and the fact that $C(z)$ surrounds both the origin and the point $z$ implies that the partial sums of $\sum c_j z^j$ are given by

$$\begin{aligned} S_k(z) &= \frac{1}{2\pi i}\int_{C(z)} \frac{f(u)}{u-z}\left[1 - \left(\frac{z}{u}\right)^{k+1}\right]du \\ &= f(z) - \frac{1}{2\pi i}\int_{C(z)} \frac{zf(u)}{u(u-z)}\left(\frac{z}{u}\right)^k du. \end{aligned} \tag{4.52}$$

The $H(\chi)$ transform of $\sum c_j z^j$ is accordingly given by

$$\sigma_n(z) = f(z) - (2\pi i)^{-1}R_n(z) \tag{4.53}$$

where

$$\begin{aligned} R_n(z) &= \int_0^1 d\chi(t)\int_{C(z)} \frac{zf(u)}{u(u-z)}\sum_{k=0}^{n} C_{n,k}\left(\frac{tz}{u}\right)^k (1-t)^{n-k}du \\ &= \int_0^1 d\chi(t)\int_{C(z)} \frac{zf(u)}{u(u-z)}\left[1 + t\left(\frac{z}{u} - 1\right)\right]^n du. \end{aligned}$$

Hence we may complete the proof of Theorem 4.1 by showing that $R_n(z)$ converges to 0 uniformly over $\beta_0$. Since $\chi(t)$ is constant over $r < t \leqq 1$, we may assume that $\chi(t) = 1$ over $r \leqq t \leqq 1$ and replace the upper limit of integration by $r$. We then obtain

$$|R_n(z)| \leqq \int_0^r |d\chi(t)| \int_{C(z)} \frac{|z||f(u)|}{|u||u-z|}\left|1 + t\left(\frac{z}{u} - 1\right)\right|^n |du|. \tag{4.54}$$

If $c_6$ denotes the least upper bound of $|z|$ for $z \in \beta_0$, then

$$\frac{|z||f(u)|}{|u||u-z|} \leqq \frac{c_6 c_5}{c_2 c_4} = c_7 \tag{4.55}$$

where $c_7$ is the constant, independent of $z$ and $u$, defined by the last equality. Let $c_8$ be the least upper bound of the circumferences of the circles $C(z)$; it is obviously finite since the circles lie in the bounded set $W(h_1)$. Let $c_9 = c_7 c_8$. With the aid of (4.34) and (4.35) we obtain, when $0 < \delta < r$

$$|R_n(z)| \leqq c_9 \int_0^\delta |d\chi(t)| + c_9 \int_\delta^r \theta_1^n |d\chi(t)|. \tag{4.6}$$

Let $\epsilon > 0$, and fix $\delta$ such that $0 < \delta < r$ and the first term on the right is less than $\epsilon/2$. Then choose $N$ such that the last term is less than $\epsilon/2$ when $n \geqq N$.

We then have $|R_n(z)| < \epsilon$ when $z \in \beta_0$ and $n \geqq N$. This completes the proof of Theorem 4.1.

Throughout this section, we have considered power series $\sum c_n z^n$ for which the radius of convergence $R$ is finite. In case $R = \infty$ the series is uniformly summable, to the entire function $f(z)$ determined by the series, over each bounded set $E$ by each regular transformation of the form

$$\sigma(t) = \sum_{k=0}^{\infty} a_k(t) s_k \tag{4.7}$$

and hence in particular by each regular transformation of the form $H(\chi)$. To prove this, we note first that if $s_n(z)$, $n = 0, 1, 2, \cdots$, is the sequence of partial sums of $\sum c_n z^n$, then the sequence converges uniformly over $E$ to $f(z)$ and the sequence is uniformly bounded over $E$. Hence (see Agnew [1, Theorem 7.21], and Agnew [2]) the sequence $s_n(z)$ and the series $\sum c_n z^n$ must be uniformly summable to $f(z)$ over $E$.

5. **Other methods of summability.** Let $\sum c_n z^n$ be a power series having a finite positive radius of convergence, and let $F$ be a closed set interior to the Borel polygon $B$. Then $r$ exists such that $0 < r \leqq 1$ and $F \subset B(r)$, and, by Theorem 4.1, $\sum c_n z^n$ is uniformly summable $E_r$ over $F$ to $f(z)$. Let $\sigma_n^{(r)}(z)$, $n = 0, 1, \cdots$, denote the $E_r$ transform of $\sum c_n z^n$. Since $f(z)$ and the functions $\sigma_1^{(r)}(z)$, $\sigma_2^{(r)}(z)$, $\cdots$ are each bounded over $F$, it follows that the sequence $\sigma_n^{(r)}(z)$ is uniformly bounded over $F$.

Let $G$ be a method of summability of the form (4.7) which includes $E_r$. For example, $G$ may be the Borel exponential method (Hurwitz [12, p. 27]) or the LeRoy method (Morse [15, p. 281]). Let $G$ be such that, for each $t$,

$$\sum_{k=0}^{\infty} a_k(t) z^k$$

converges for all complex values of $z$. This requirement, which is automatically satisfied when $G$ is of finite reference, is also satisfied when $G$ is either the Borel exponential method or the LeRoy method. Since the inverse of $E_r$ is $E_{1/r}$ (Hurwitz [12]), the $G$ transform of $s_n(z)$ may be written

$$\sigma(t, z) = \sum_{k=0}^{\infty} \sum_{p=0}^{k} a_k(t) C_{k,p} \frac{1}{r^p}\left(1 - \frac{1}{r}\right)^{k-p} \sigma_p^{(r)}(z). \tag{5.1}$$

Our hypotheses imply that, for each $t$, the series in the right member of (5.1) is absolutely convergent. Hence (5.1) can be written in the form

$$\bar{\sigma}(t) = \sum_{p=0}^{\infty} \left[ \sum_{k=p}^{\infty} a_k(t) C_{k,p} \frac{1}{r^p}\left(1 - \frac{1}{r}\right)^{k-p} \right] \bar{s}_p \tag{5.2}$$

where $\bar{\sigma}(t) = \sigma(t, z)$ and $\bar{s}_p = \sigma_p^{(r)}(z)$. The hypothesis that $G$ includes $E_r$ im-

plies that (5.2) is a regular transformation of the form (4.7). Hence it follows, as at the end of the last section, that $\sigma(t, z)$ must converge to $f(z)$ uniformly over $F$, that is, $\sum c_n z^n$ is summable $G$ to $f(z)$ uniformly over $F$. In particular, *a power series* $\sum c_n z^n$ *is summable by the Borel exponential method uniformly over each bounded closed set inside the Borel polygon.* This particular result may be known, but the author is unable to give a reference. Borel [4] and Phragmén [16] have shown that $\sum c_n z^n$ is summable, by the exponential method, at each point inside the Borel polygon and non-summable at each point outside the polygon; and Doetsch [5] has discussed summability on the boundary of the polygon.

6. **Collective Hausdorff summability** $\mathcal{H}$. It was shown by Hurwitz and Silverman [13] that a transformation of the form

$$\sigma_n = \sum_{k=0}^{\infty} a_{nk} s_k \tag{6.11}$$

commutes with the $C_1$ transformation

$$\sigma_n = \frac{1}{n+1} \sum_{k=0}^{n+1} s_k \tag{6.12}$$

if and only if it has the form

$$\sigma_n = \sum_{k=0}^{n} C_{n,k} \left\{ \sum_{j=k}^{n} (-1)^{j-k} C_m - k, j - k\lambda_j \right\} s_k \tag{6.13}$$

where $\lambda_0, \lambda_1, \cdots$ is a sequence of complex constants. It was shown by Hausdorff [9] that such a transformation is regular if and only if the sequence $\lambda_0, \lambda_1, \cdots$ is the moment sequence

$$\lambda_n = \int_0^1 t^n d\chi(t) \tag{6.14}$$

of a function $\chi(t)$, having bounded variation over $0 \leqq t \leqq 1$, such that $\chi(0+) = \chi(0) = 0$ and $\chi(1) = 1$. When (6.14) holds, the transformation (6.13) takes the form $H(\chi)$. As was shown both by Hurwitz and Silverman and by Hausdorff, these regular transformations commute with each other as well as with $C_1$ and hence constitute a system of consistent methods of summability. By this we mean that if a series or sequence is summable by two different regular methods of the form $H(\chi)$, then the two values assigned must be equal.

This circumstance makes possible the following definition of a method $\mathcal{H}$ of summability which makes use of the collection of regular methods $H(\chi)$ and which may be called the *collective Hausdorff method*. Let a series $\sum u_n$ be called *summable* $\mathcal{H}$ to the value $\sigma$ if it is summable to the value $\sigma$ by at least one regular method $H(\chi)$.

The method $\mathfrak{H}$ is obviously regular. It is also linear; by this we mean that if two series $\sum u_n$ and $\sum v_n$ with partial sums $s_n$ and $t_n$ are summable $\mathfrak{H}$ to $U$ and $V$, respectively, and if $\alpha$ and $\beta$ are constants, then $\sum(\alpha u_n+\beta v_n)$ is summable $\mathfrak{H}$ to $\alpha U+\beta V$. This follows from Hurwitz and Silverman [13, Theorem 3].

It is clear from Theorem 1.1 that the method $\mathfrak{H}$ is stronger than any one method of the form $H(\chi)$; if $H(\chi)$ is regular, then $\sum z^n$ is summable $H(\chi)$ only inside and perhaps at some of the points on the boundary of the appropriate circle of summability, but $\sum z^n$ is summable $\mathfrak{H}$ for each $z$ in the half-plane $\mathfrak{R}z<1$. It would be interesting to know whether there is a regular method of summability, based on a single sequence-to-sequence or sequence-to-function transformation of the familiar type, which includes $\mathfrak{H}$; perhaps there is one which is equivalent to $\mathfrak{H}$. The Borel exponential method $B_1$ evaluates $\sum z^n$ when $\mathfrak{R}(z)<1$; but $B_1$ does not include the method $C_1$ determined by $\chi(t)=t$ and hence $B_1$ does not include $\mathfrak{H}$. The LeRoy method $L$ includes the regular Euler methods $E_r$ (Morse [15]) and the regular Cesàro methods $C_r$ (Garabedian [6]); but, if one may judge from the difficulty in showing that $L$ includes $C_r$, it must be difficult to determine the extent to which $L$ includes Hausdorff methods. That $L$ does not include $\mathfrak{H}$ is a corollary of Theorem 8.4.

It follows from Theorem 1 that the series $\sum z^n$ is non-summable $\mathfrak{H}$ for each $z$ for which $\mathfrak{R}z>1$. This result creates a strong presumption that $\mathfrak{H}$ is ineffective outside the Borel polygon of a power series $\sum a_n z^n$. If a point $z_0$ lies outside the Borel polygon of a series $\sum a_n z^n$, then $z_0$ lies outside the circle of convergence and accordingly the series $\sum a_n z_0^n$ has unbounded partial sums.

We now show existence of series with bounded partial sums $s_n$ which are not summable $\mathfrak{H}$. The series are gap series with large gaps. Let $n_0, n_1, n_2, \cdots$ be a sequence of integers for which $0=n_0<n_1<n_2<\cdots$ and

$$n_{p+1}/n_p \to \infty \tag{6.2}$$

as $p\to\infty$. Let $b_1, b_2, \cdots$ be a bounded divergent sequence of complex numbers, and let $\sum u_n$ be the series whose partial sums $s_0, s_1, \cdots$ are defined by the formulas

$$s_k = b_p, \qquad n_{p-1} \leqq k \leqq n_p;\ p = 1, 2, \cdots. \tag{6.21}$$

Let $H(\chi)$ be regular; we show that $\sum u_n$ is non-summable $\mathfrak{H}$ by showing that it is non-summable $H(\chi)$. Letting $\sigma_n$ denote the $H(\chi)$ transform of $\sum u_n$, and setting $m_p=n_p-1$, we find that when $p>1$

$$\sigma(m_p) = \int_0^1 \sum_{k=0}^{m_p} C_{m_p,k} t^k (1-t)^{m_p-k} s_k d\chi(t)$$

and

$$(6.3)\qquad \begin{aligned}\sigma(m_p) - s(m_p) &= \int_0^1 \sum_{k=0}^{m_p} C_{m_p,k} t^k (1-t)^{m_p-k}[s_k - s(m_p)]d\chi(t)\\ &= \int_0^1 \sum_{k=0}^{m_p-1} C_{m_p,k} t^k (1-t)^{m_p-k}[s_k - s(m_p)]d\chi(t)\end{aligned}$$

so that, where $M$ is a constant for which $|b_k| \leqq M$ for each $k=0, 1, 2, \cdots$,

$$|\sigma(m_p) - s(m_p)| \leqq 2M \int_0^1 \sum_{k=0}^{m_p-1} C_{m_p,k} t^k (1-t)^{m_p-k} |d\chi(t)|.$$

Let $\epsilon > 0$. Choose $\delta > 0$ such that

$$2M \int_0^\delta |d\chi(t)| < \epsilon/2.$$

Let $0 < \theta < \delta$ and choose an index $P$ such that $m_{p-1} < \theta m_p$ when $p \geqq P$. Then when $p \geqq P$

$$|\sigma(m_p) - s(m_p)| < \epsilon/2 + 2M \int_\delta^1 \sum_{k<\theta m_p} C_{m_p,k} t^k (1-t)^{m_p-k} |d\chi(t)|.$$

If $L_1, L_2, \cdots$ is a sequence of positive constants for which $L_n \to \infty$ and and $L_n/n \to 0$, say $L_n = n^{1/2}$, we can use the elementary inequality (see, for example, Hausdorff [9, p. 104])

$$\sum_{k<k(n,t)} C_{n,k} t^k (1-t)^{n-k} \leqq \frac{1}{4L_n},$$

where $k(n, t) = n\{t - (L_n/n)^{1/2}\}$, to obtain, when $p$ is sufficiently great,

$$|\sigma(m_p) - s(m_p)| < \frac{\epsilon}{2} + \frac{2M}{4L(m_p)} \int_\delta^1 |d\chi(t)|.$$

Since the last term is less than $\epsilon/2$ when $p$ is sufficiently great, and since $s(m_p) = b_p$, this implies that

$$(6.4)\qquad \lim_{p\to\infty} |\sigma(m_p) - b_p| = 0.$$

Thus divergence of the sequence $b_n$ implies that of the sequence $\sigma_n$ and accordingly $\sum u_n$ is non-summable $H(\chi)$.

7. **A Tauberian theorem for Hausdorff methods.** It is easy to amplify the work of the preceding paragraph to establish the following Tauberian gap theorem in which there is no order condition placed upon the non-vanishing terms or partial sums of the series.

THEOREM 7.1. *If $n_0, n_1, n_2, \cdots$ is a sequence of integers for which $0 = n_0 < n_1 < n_2 < \cdots$ and*

$$n_{p+1}/n_p \to \infty$$

*as $p\to\infty$, if $\sum u_n$ is a series for which*

$$u_n = 0, \qquad n \neq n_0, n_1, n_2, \cdots,$$

*and if $\sum u_n$ is summable $\mathcal{H}$, then $\sum u_n$ is convergent*(8).

The hypothesis implies existence of a sequence $b_1, b_2, \cdots$ such that the partial sums $s_0, s_1, \cdots$ of $\sum u_n$ satisfy (6.21). Let $H(\chi)$ be a regular Hausdorff method by which $\sum u_n$ is summable. In case the sequence $b_n$ is bounded, we can proceed exactly as above to obtain (6.4). Convergence of $\sigma_n$ then implies that of $b_n$ and hence that of $\sum u_n$. It remains for us to show that the sequence $b_k$ must be bounded. For each $p=1, 2, \cdots$ let $M_p$ denote the maximum value of $|s_k|$ when $0\leqq k\leqq m_p$ where, as above, $m_p=n_p-1$. Then we can obtain (6.3) and conclude that

$$\left| \sigma(m_p) - s(m_p) \right| \leqq 2M_p \int_0^1 \sum_{k=0}^{m_p-1} C_{m_p,k} t^k (1-t)^{m_p-k} \left| d\chi(t) \right|.$$

Choose $\delta>0$ such that

$$\int_0^\delta \left| d\chi(t) \right| < 1/5$$

and then choose an index $P$ such that

$$\int_\delta^1 \sum_{k=0}^{m_p-1} C_{m_p,k} t^k (1-t)^{m_p-k} \left| d\chi(t) \right| < 1/5, \qquad p \geqq P.$$

Then

$$\left| \sigma(m_p) - s(m_p) \right| \leqq (4/5) M_p, \qquad p \geqq P.$$

If the sequence $b_k$ is unbounded, then $M_p\to\infty$ as $p\to\infty$ and there is an infinite set of indices $p\geqq P$ for which $|s(m_p)|=M_p$. For such values of $p$, $|\sigma(m_p)|\geqq M_p/5$. This is inconsistent with the hypothesis that $\sum u_n$ is summable $H(\chi)$; hence the sequence $b_k$ must be bounded and Theorem 7.1 is proved.

8. **Criteria involving zeros of moment functions.** Let $H(\chi)$ be regular, and let

$$\mu(z) = \int_0^1 t^z d\chi(t), \qquad z \geqq 0,$$

be the moment function determined by $\chi(t)$. The function $\mu(z)$ is continuous in the closed half-plane $\mathfrak{R}z\geqq 0$ and is analytic in the open half-plane $\mathfrak{R}z>0$.

(8) For a Tauberian theorem involving a subclass of the regular Hausdorff methods, see Pitt [17, pp. 280–284] and Agnew [3].

Several investigations have shown that the zeros of $\mu(z)$ play a fundamental role in the theory of transformations of the form $H(\chi)$. This applies, in particular, to the relation between $H(\chi)$ and generalized Abel summability $A^*$ used by Silverman and Tamarkin [20]. Let a series $\sum u_n$ with partial sums $s_n$ be called summable $A^*$ to $\sigma^*$ if the series

$$\sum_{k=0}^{\infty} (1 - z)z^k s_k$$

has a positive radius of convergence and defines, by analytic extension along radial lines from the origin, a function $\sigma^*(z)$ such that $\sigma^*(z)$ exists when $0<z<1$ and $\sigma^*(z)\to\sigma^*$ as $z\to 1$ over the set $0<z<1$. It was shown by Silverman and Tamarkin [20] that if $\mu(z)$ has a zero with real part positive, then $A^*$ does not include $H(\chi)$; and that if $\mu(z)$ has a zero with real part 0 and a certain supplementary condition on $\chi(t)$ is satisfied, then again $A^*$ does not include $H(\chi)$. In this section we obtain some results involving zeros of $\mu(z)$ on the critical line $\mathcal{R}z=0$; in particular we remove the supplementary condition of Silverman and Tamarkin by proving the following theorem.

THEOREM 8.1. *If $\chi(t)$ is regular and if $\mu(z)$ has a zero $q$ for which $\mathcal{R}q=0$, then $A^*$ does not include $H(\chi)$.*

Our results are obtained from consideration of the sequence

$$s_k^{(q)} = k^q = e^{q \log k}$$

in which $q$ is a complex number with $\mathcal{R}q \geqq 0$ and $q\neq 0$. Let $H(\chi)$ be regular, and let $\sigma_n^{(q)}$ denote the $H(\chi)$ transform of the sequence $s_n^{(q)}$. Then

$$\sigma_n^{(q)}/n^q = \int_0^1 f_n(t)d\chi(t) \tag{8.11}$$

where

$$f_n(t) = \sum_{k=0}^{n} C_{n,k}t^k(1 - t)^{n-k}f(k/n) \tag{8.12}$$

and

$$f(t) = t^q. \tag{8.13}$$

The function $f(t)$ is bounded, is continuous over $\delta \leqq t \leqq 1$ for each $\delta>0$, and is continuous at $t=0$ if and only if $\mathcal{R}q>0$. For each $n$, the function $f_n(t)$ is (see Hausdorff, [9, p. 104]) the $n$th Bernstein polynomial determined by $f(t)$, and $f_n(t)$ converges uniformly to $f(t)$ over each interval $0<\delta\leqq t\leqq 1$. Let $\epsilon>0$. Choose $\delta>0$ such that

$$\int_0^\delta |d\chi(t)| < \epsilon/2. \tag{8.14}$$

Then, since $|f_n(t)| \leqq 1$ and $|f(t)| \leqq 1$ when $0 \leqq t \leqq 1$ and $n = 0, 1, 2, \cdots,$

$$\begin{aligned}
&\left| \int_0^1 f_n(t)d\chi(t) - \int_0^1 f(t)d\chi(t) \right| \\
(8.15) \qquad &\leqq \int_0^\delta |f_n(t) - f(t)| \, |d\chi(t)| + \int_\delta^1 |f_n(t) - f(t)| \, |d\chi(t)| \\
&\leqq \epsilon + \left[ \max_{\delta \leqq t \leqq 1} |f_n(t) - f(t)| \right] \int_\delta^1 |d\chi(t)|.
\end{aligned}$$

This implies that, as $n \to \infty$, the superior limit of the first member of (8.15) is less than or equal to $\epsilon$ and hence 0. Hence we can let $n \to \infty$ in (8.11) to obtain

$$(8.16) \qquad \lim_{n \to \infty} \frac{\sigma_n^{(q)}}{n^q} = \int_0^1 t^q d\chi(t) = \mu(q).$$

Suppose now that $\Re q = 0$ but $q \neq 0$, say $q = iy$ where $y$ is real and $y \neq 0$. Then the sequence

$$s_k^{(q)} = k^q = e^{iy \log k}$$

is a sequence of points, on the unit circle, having the unit circle for its set of limit points. If $\mu(q) = 0$, then (8.16) implies that $\sigma_n^{(q)} \to 0$; but if $\mu(q) \neq 0$, then $\sigma_n^{(q)}$ is a divergent sequence whose limit points constitute a circle with radius $|\mu(q)|$. Thus we obtain the following theorem which has several applications.

THEOREM 8.2. *If $H(\chi)$ is regular and $\Re q = 0$ but $q \neq 0$, then the sequence $s_k^{(q)} \equiv k^q$ is summable $H(\chi)$ if and only if $\mu(q) = 0$.*

We now use Theorem 8.2 to prove Theorem 8.1. Let $H(\chi)$ be regular and let $q$ be a zero of $\mu(z)$ for which $\Re(q) = 0$. Since regularity of $H(\chi)$ implies that $\mu(0) = 1$, we have $q \neq 0$. Hence, by Theorem 8.2, the sequence $k^q$ is summable $H(\chi)$. The sequence $k^q$ is the sequence of partial sums of a divergent series $\sum u_n$ for which $n|u_n|$ is bounded; hence an elementary Tauberian theorem implies that the sequence $k^q$ is not summable by the ordinary Abel method $A$. Since the series $\sum k^q z^k$ has radius of convergence 1, it follows that the sequence $k^q$ is also not summable by the generalized Abel method $A^*$. Existence of the sequence summable $H(\chi)$ but not $A^*$ shows that $A^*$ does not include $H(\chi)$ and establishes Theorem 8.1.

Another consequence of Theorem 8.2 is set forth in the following theorem in which it is not assumed that the transformation $H(\chi_1)$ has an inverse.

THEOREM 8.3. *If $H(\chi_2)$ and $H(\chi_1)$ are two regular Hausdorff methods such that $H(\chi_2)$ includes $H(\chi_1)$, then each zero of the moment function $\mu_1(z)$ of $\chi_1(t)$ with real part 0 is also a zero of the moment function $\mu_2(z)$ of $\chi_2(t)$.*

To prove this theorem, let $q$ be a zero of $\mu_1(z)$ for which $\mathfrak{R}q=0$. Then $q\neq 0$, and Theorem 8.2 implies that the sequence $k^q$ is summable $H(\chi_1)$. Hence $k^q$ is also summable $H(\chi_2)$ and Theorem 8.2 implies that $\mu_2(q)=0$.

In case $H(\chi_1)$ and $H(\chi_2)$ satisfy the hypothesis of Theorem 8.3 and $H(\chi_1)$ has an inverse $[H(\chi_1)]^{-1}$, then the transformation $H(\chi_2)[H(\chi_1)]^{-1}$ is a regular Hausdorff transformation $H(\chi_3)$ such that $H(\chi_2)=H(\chi_3)H(\chi_1)$, and the moment function $\mu_3(z)$ of $\chi_3(t)$ is such that

$$(8.31)\qquad \mu_2(z) = \mu_3(z)\mu_1(z), \qquad z \geqq 0.$$

This result was established by Hurwitz and Silverman [13] for the case in which $\mu_1(z)$ and $\mu_2(z)$ are analytic at $\infty$ and in a half-plane $\mathfrak{R}z>-\alpha$ for some $\alpha>0$; the extension to the more general regular transformations was made by Hille and Tamarkin [11] and by Garabedian, Hille, and Wall [7]. It thus appears that if $H(\chi_1)$ has an inverse, then the conclusion of Theorem 8.3 is a trivial consequence of (8.31). It would be interesting to know whether Theorem 8.3 could be strengthened by establishing existence of a regular moment function $\mu_3(z)$ satisfying (8.31) even when $H(\chi_1)$ does not have an inverse.

The sequences $k^q$, in which $\mathfrak{R}q=0$, $q\neq 0$, are not the only sequences of which the question of summability $H(\chi)$ is settled by vanishing or nonvanishing of a single moment of $\chi(t)$. Let $p$ be a positive integer and let

$$s_k^{(p)} = k!/(k-p)!, \qquad k = 0, 1, 2, \cdots,$$

where $1/(k-p)!$ is interpreted to be 0 when $k=0, 1, \cdots, p-1$. Let $H(\chi)$ be regular, and let $\sigma_n^{(p)}$ denote the $H(\chi)$ transform of $s_n^{(p)}$. Then

$$\begin{aligned}\sum_{k=0}^{n} C_{n,k}t^k(1-t)^{n-k}s_k^{(p)} &= \frac{n!}{(n-p)!}t^p\sum_{k=0}^{n-p} C_{n-p,k}t^k(1-t)^{n-p-k}\\ &= \frac{n!}{(n-p)!}t^p[t+1-t]^{n-p} = \frac{n!}{(n-p)!}t^p\end{aligned}$$

so that

$$\sigma_n^{(p)} = \frac{n!}{(n-p)!}\int_0^1 t^p d\chi(t) = \frac{n!}{(n-p)!}\mu(p);$$

this formula was obtained by Silverman [19] by another method. It is a consequence of this formula that the sequence $n!/(n-p)!$, in which $p$ is a positive integer, is summable by a regular Hausdorff method $H(\chi)$ if and only if $\mu(p)=0$, $\mu(z)$ being the moment function of $\chi(t)$. This implies that if $H(\chi_2)$ and $H(\chi_1)$ are regular Hausdorff methods with moment functions $\mu_2(z)$ and $\mu_1(z)$, and if $H(\chi_2)$ includes $H(\chi_1)$, then each positive integer zero of $\mu_1(z)$ must be a zero of $\mu_2(z)$.

We are now in a position to prove the following theorem which, in particular, implies that the LeRoy method $L$ does not include $\mathfrak{H}$.

THEOREM 8.4. *No totally regular method of summability can include* $\mathfrak{H}$.

A method $T$ of summability is totally regular if it is regular, and if also the $T$ transform of each real sequence $s_n$ for which $s_n \to +\infty$ diverges to $+\infty$. Let $p$ be a positive integer and let $H(\chi)$ be a regular Hausdorff method such that the moment function $\mu(z)$ of $\chi(t)$ vanishes when $z=p$. Then the sequence $n!/(n-p)!$ is summable $H(\chi)$ to 0. Hence the sequence is summable $\mathfrak{H}$ to 0. But when $T$ is totally regular, the sequence is not summable $T$. This proves Theorem 8.4.

## REFERENCES

1. R. P. Agnew, *The behavior of bounds and oscillations of sequences of functions under regular transformations*, these Transactions, vol. 32 (1930), pp. 669–708.

2. ———, *The effects of general regular transformations on oscillations of sequences of functions*, these Transactions, vol. 33 (1931), pp. 411–424.

3. ———, *Some remarks on a paper entitled "General Tauberian theorems,"* Journal of the London Mathematical Society, vol. 15 (1940), pp. 242–246.

4. E. Borel, *Leçons sur les Fonctions Entières*, Paris, 1900.

5. G. Doetsch, *Über die Summabilität von Potenzreihen auf dem Rande des Borelschen Summabilitätspolygons*, Mathematische Annalen, vol. 84 (1921), pp. 245–251.

6. H. L. Garabedian, *On the relation between certain methods of summability*, Annals of Mathematics, (2), vol. 32 (1931), pp. 83–106.

7. H. L. Garabedian, E. Hille, and H. S. Wall, *Formulations of the Hausdorff inclusion problem*, Duke Mathematical Journal, vol. 8 (1941), pp. 193–213.

8. H. L. Garabedian and H. S. Wall, *Hausdorff methods of summation and continued fractions*, these Transactions, vol. 48 (1940), pp. 185–207.

9. F. Hausdorff, *Summationsmethoden und Momentfolgen*, I and II, Mathematische Zeitschrift, vol. 9 (1921), pp. 74–109 and 280–299.

10. E. Hille, *Essai d'une bibliographie de la representation analytique d'une fonction monogène*, Acta Mathematica, vol. 52 (1929), pp. 1–80.

11. E. Hille and J. D. Tamarkin, *Questions of relative inclusion in the domain of Hausdorff means*, Proceedings of the National Academy of Sciences, vol. 19 (1933), pp. 573–577.

12. W. A. Hurwitz, *Report on topics in the theory of divergent series*, Bulletin of the American Mathematical Society, vol. 28 (1922), pp. 17–36.

13. W. A. Hurwitz and L. L. Silverman, *On the consistency and equivalence of certain definitions of summability*, these Transactions, vol. 18 (1917), pp. 1–20.

14. K. Knopp, *Über das Eulersche Summierungsverfahren*, I and II, Mathematische Zeitschrift, vol. 15 (1922), pp. 226–253 and vol. 16 (1923), pp. 125–156.

15. D. S. Morse, *Relative inclusiveness of certain definitions of summability*, American Journal of Mathematics, vol. 45 (1923), pp. 259–285.

16. E. Phragmén, *Sur le domaine de convergence de l'intégrale infinie* $\int_0^\infty F(ax)e^{-a}da$, Comptes Rendus de l'Académie des Sciences, Paris, vol. 132 (1901), pp. 1396–1399.

17. H. R. Pitt, *General Tauberian theorems*, Proceedings of the London Mathematical Society, (2), vol. 44 (1938), pp. 243–288.

18. H. Rademacher, *Über den Konvergenzbereich der Eulerschen Reihentransformation*, Sitzungsberichte der Berliner Mathematischen Gesellschaft, vol. 21 (1922), pp. 16–24.

19. L. L. Silverman, *On the omission of terms in certain summable series*, Recueil Mathématique, vol. 33 (1926), pp. 375–384.

20. L. L. Silverman and J. D. Tamarkin, *On the generalization of Abel's theorem for certain definitions of summability*, Mathematische Zeitschrift, vol. 29 (1928), pp. 161–170.

CORNELL UNIVERSITY,
ITHACA, N. Y.

# THE SPECTRUM OF LINEAR TRANSFORMATIONS

BY

EDGAR R. LORCH

## I. Introduction

The determination of the structure of a linear transformation $T$ which maps a complex vector space $\mathfrak{B}$ into itself is naturally a primary object of the theory of linear transformations. The first stage of such a study centers about the question of reducibility. The concept may be introduced in various ways. In one formulation it is required to find all projections (bounded idempotent transformations) $P$ which commute with $T$. At a more incisive level, one seeks those pairs of manifolds $\mathfrak{M}$ and $\mathfrak{N}$ (not necessarily disjoint) which "split" the space and are transformed into themselves by $T$. For transformations in general vector spaces the known results all refer to special types such as the completely continuous or the weakly almost periodic. This paper will deal almost exclusively with the first type of reducibility of the general bounded transformation. The boundedness of $T$ is assumed for convenience; in case merely its closure is hypothesized the salient features of the theory are still valid.

The results are all based on one method, that of a contour integral of the resolvent of $T$. They seem to exhaust the possibilities for this particular tool. Means for cracking the spectrum directly would undoubtedly have to be of a much more delicate nature. The fundamental projection is the integral

$$P = \frac{1}{2\pi i}\int_C \frac{d\zeta}{\zeta I - T}$$

evaluated over a simple closed contour lying entirely in the resolvent set of $T$(1). An integral bearing suggestive resemblance to this one is used by E. Hille in an analysis of semi-groups of linear transformations(2). The range of $P$ consists entirely of those elements in $\mathfrak{B}$ associated with the spectral val-

Presented to the Society, May 2, 1941; received by the editors March 24, 1941, and, in revised form, July 16, 1941.

(1) This transformation is well known in the theory of finite matrices. It was introduced by Frobenius, *Über die schief Invariante einer bilinearen oder quadratischen Form*, Crelles Journal, vol. 86 (1879), pp. 44–71. The construction of $P$ and the demonstration of its properties in this case are quite simple and are carried out by methods characteristic of that theory.

(2) E. Hille, *Notes on linear transformations*. II. *Analyticity of semi-groups*, Annals of Mathematics, (2), vol. 40 (1939), pp. 1–47. As a principal problem of this paper is one in interpolation, Professor Hille uses an integral of the type $T^\alpha = (2\pi i)^{-1}\int_C \zeta^\alpha(\zeta I - T)^{-1}d\zeta$ where $\alpha$ is a complex power and $C$ represents a simple closed curve essentially containing the entire spectrum of $T$ in its interior. Thus the emphasis here is on the factor $\zeta^\alpha$. In our case, on the contrary, the choice of the curve $C$ is paramount. Added November 12, 1941.

ues lying within the curve $C$. Similarly the spectrum of $T$ over the range of $I-P$ lies outside of $C$. If one associates the projection $P$ with the point set interior to $C$ one has an instance of a mapping of so-called spectral sets upon projections. It is shown that this mapping is a homomorphism. A spectral set is any set obtained from the interiors of curves $C$ by the operations of complementation, addition, and intersection performed a finite number of times. The attempt to extend the homomorphism so as to admit of denumerable addition and intersection has to be abandoned even if the space $\mathfrak{B}$ is required to be reflexive and the more general type of reducibility (vide supra) is considered.

## II. The fundamental projections

**Notations and definitions.** The notations used are these: Complex numbers are denoted by $\alpha, \beta, \gamma, \zeta, \lambda, \xi$, and so on; elements of the space $\mathfrak{B}$ by $f, g, h$, and so on; linear transformations of $\mathfrak{B}$ into a subset of itself by $T, S, I$ (the identity), 0 (the zero), $P$, and so on.

The underlying space $\mathfrak{B}$ is a normed linear complex vector space, that is, a complex Banach space. The transformations $T$ are assumed to be linear or distributive, $T(\alpha f+\beta g)=\alpha Tf+\beta Tg$. Unless explicitly stated, $T$ is bounded, $\|Tf\| \leqq K\|f\|$, for arbitrary $f\in\mathfrak{B}$; the bound of $T$ is denoted by $|T|$.

A number $\lambda$ is said to belong to the *resolvent set* $\mathfrak{R}$ of $T$ if the transformation $T-\lambda I$ maps $\mathfrak{B}$ upon itself in a one-to-one manner. It results from this fact that $(T-\lambda I)^{-1}$ is a bounded linear transformation([3]).

If for a given $\lambda$, $Tf-\lambda f=0$ for an $f\neq 0$, $\lambda$ is said to belong to the *point spectrum* of $T$. If $\lambda$ belongs neither to the point spectrum nor to the resolvent set, the transformation $T-\lambda I$ may be inverted and has a range which is either dense in $\mathfrak{B}$ (but not identical with $\mathfrak{B}$), or is not dense in $\mathfrak{B}$. In the first case $\lambda$ is said to belong to the *continuous spectrum* of $T$; in the second, $\lambda$ is said to belong to the *residual spectrum* of $T$. If $\lambda$ is in the continuous spectrum, the transformation $(T-\lambda I)^{-1}$ is unbounded and there exist elements $f_n\in\mathfrak{B}$, $n=1, 2, \cdots$, such that $\|f_n\|=1$, $\|(T-\lambda I)f_n\|\to 0$. The collection of values $\lambda$ in the point, continuous, and residual spectra is called the *spectrum* $\mathfrak{S}$ of $T$. These three classes in addition to the resolvent set are mutually exclusive and all inclusive in the complex plane.

**Some elementary theorems.** In this section some elementary theorems are stated.

THEOREM 1. *The value $\lambda$ is in the resolvent set (spectrum) of $T$ if and only if the value $\bar{\lambda}$ is in the resolvent set (spectrum) of the adjoint $\overline{T}$ of $T$. If $\lambda$ is in the point spectrum of $T$, $\bar{\lambda}$ is in the point or residual spectrum of $\overline{T}$. If $\lambda$ is in the residual spectrum of $T$, $\bar{\lambda}$ is in the point spectrum of $\overline{T}$. If $\lambda$ is in the continuous spectrum of $T$, $\bar{\lambda}$ is in the continuous or residual spectrum of $\overline{T}$. If the space $\mathfrak{B}$*

([3]) Banach, *Théorie des Opérations Linéaires*, p. 41.

*is reflexive and* $\lambda$ *is in the continuous spectrum of* $T$, *then,* $\bar{\lambda}$ *is in the continuous spectrum of* $\bar{T}$(4).

The theorem is an immediate consequence of the definition of the adjoint transformation and of some elemental properties of linear transformations.

THEOREM 2. *If* $\lambda$ *belongs to the resolvent set* $\mathfrak{R}$ *of* $T$, *then so do all* $\zeta$ *with* $|\zeta-\lambda|<1/|(T-\lambda I)^{-1}|$. *Thus* $\mathfrak{R}$ *is an open set and the spectrum* $\mathfrak{S}$ *is a closed set. The transformation* $(T-\zeta I)^{-1}$ *may be expressed "analytically" in terms of* $(T-\lambda I)^{-1}$ *by means of the formula*(5):

$$(1)\ (T-\zeta I)^{-1}=(T-\lambda I)^{-1}+(\zeta-\lambda)(T-\lambda I)^{-2}+(\zeta-\lambda)^2(T-\lambda I)^{-3}+\cdots.$$

*The bound of* $(T-\zeta I)^{-1}$ *is a continuous function of* $\zeta$.

No proof will be given.

That the resolvent set is not empty may be shown as follows: Let $\lambda$ be so chosen that $|\lambda|>|T|$. Then since $\|(T-\lambda I)f\|$ is bounded away from zero, $\lambda$ is neither in the point nor in the continuous spectrum. By Theorem 1 and the fact that $|T|=|\bar{T}|$, $\lambda$ is not in the residual spectrum. Thus $\lambda\in\mathfrak{R}$. The fact that $\mathfrak{S}$ is not empty will be shown later.

THEOREM 3. *If* $\lambda$ *and* $\mu$ *are any two values in the resolvent set* $\mathfrak{R}$ *of* $T$, *then*(6)

$$(2)\qquad (\lambda I-T)^{-1}-(\mu I-T)^{-1}=(\mu-\lambda)(\lambda I-T)^{-1}(\mu I-T)^{-1}.$$

In proof, multiply both sides of (2) by $(\lambda I-T)(\mu I-T)$.

Let $T(\zeta)$ be a function defined over some set of complex numbers $\zeta$ and whose values are bounded linear transformations of $\mathfrak{B}$ into itself. For such a transformation depending continuously upon a parameter will be defined a curvilinear integral along a given rectifiable curve $C$. The function $T(\zeta)$ is said to be continuous if it is continuous in the uniform topology, that is, $|T(\zeta_1)-T(\zeta_2)|$ is small with $|\zeta_1-\zeta_2|$.

LEMMA. *Let* $C$ *be a rectifiable curve in the complex plane,* $\zeta=\phi(t)$, $0\leqq t\leqq 1$, *and* $T(\lambda)$ *be a transformation depending continuously upon a parameter* $\lambda$ *and defined on* $C$. *Then the Riemann integral*

$$(3)\qquad A=\int_C T(\zeta)d\zeta$$

(4) Some of these statements may be found established for Hilbert space by Stone, *Linear Transformations in Hilbert Space and Their Applications to Analysis*, American Mathematical Society Colloquium Publications, vol. 15, chap. 4.

(5) This is the well known Neumann expansion of the resolvent. The fact that this formula holds for a general transformation in a Banach space has been noted by A. E. Taylor, *The resolvent of a closed transformation*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 70–74. In the same paper one will find our Theorem 3.

(6) This is the well known functional equation for the resolvent.

*exists and defines a bounded linear transformation. The bound of $A$ satisfies $|A| \leqq \max_C |T(\zeta)| \cdot l$ where $l$ is the length of $C$. If the transformations $T(\zeta)$ are commutative, then $A$ is commutative with $T(\zeta)$.*

The integral is defined in the classic way. Let $t_0=0<t_1< \cdots <t_n=1$ be a subdivision of the unit interval. Let $t_i'$ with $t_{i-1} \leqq t_i' \leqq t_i$, $i=1, \cdots, n$ be $n$ points, one in each subinterval. Let $\zeta_i' = \phi(t_i')$ and $\Delta\zeta_i = \phi(t_i) - \phi(t_{i-1})$. Then an approximating sum to $A$ is $\sum_{i=1}^{n} T(\zeta_i')\Delta\zeta_i$. The proof of the convergence is omitted. The remaining statements of the theorem are clear.

**The fundamental projections.** In this section will be introduced the projections which underlie the entire investigation.

THEOREM 4. *Let $C$ be a simple closed rectifiable curve lying entirely within the resolvent set $\mathfrak{R}$ of $T$. Then the contour integral*

$$P = \frac{1}{2\pi i}\int_C \frac{d\zeta}{\zeta I - T} \tag{4}$$

*exists and represents a bounded linear transformation commutative with $T$, $PT=TP$. The transformation $P$ is unchanged if the curve $C$ is continuously deformed into a curve $C'$, providing only that the deformation is effected without going outside of the resolvent set.*

It is hardly necessary to state that the transformation $\zeta I - T$ appearing in the denominator of (4) represents $(\zeta I - T)^{-1}$. By Theorem 2, $|(\zeta I - T)^{-1}|$ is a continuous function of $\zeta$; hence by the lemma, the integral (4) exists defining a bounded transformation $P$. Note that the direction of integration along $C$ is counterclockwise.

Let $\lambda$ be in $\mathfrak{R}$ and $K$ be a rectifiable closed curve simple or not, lying entirely within the $\zeta$-circle $|\zeta - \lambda| = r < 1/|(T-\lambda I)^{-1}|$. The integral

$$\int_K \frac{d\zeta}{\zeta I - T} \tag{5}$$

is computed with the help of series (1); termwise integration shows that its value is zero. Now the curve $C$ may be deformed into the curve $C'$ with the help of a finite number of curves of the type $K$. Thus the integral (4) evaluated over $C$ is identical with that integral over $C'$.

THEOREM 5. *Let $C$ and $C'$ be two simple closed rectifiable curves lying entirely within the resolvent set $\mathfrak{R}$ of $T$. Let $P$ and $P'$ be the integrals* (4) *associated with $C$ and $C'$, respectively. If the curve $C$ lies entirely within the curve $C'$ then*

$$PP' = P'P = P. \tag{6}$$

*If the curves $C$ and $C'$ lie each exterior to the other, then*

$$PP' = P'P = 0. \tag{7}$$

That $PP'=P'P$ is clear from the integral definition of both $P$ and $P'$ which brands them as "functions of $T$."

By virtue of Theorem 3, one may write for an arbitrary pair of curves $C$ and $C'$

$$PP' = \left(\frac{1}{2\pi i}\right)^2 \int_C \int_{C'} \frac{1}{\zeta I - T} \cdot \frac{1}{\xi I - T} d\zeta d\xi$$
$$= \left(\frac{1}{2\pi i}\right)^2 \int_C \int_{C'} \left(\frac{1}{\zeta I - T} - \frac{1}{\xi I - T}\right) \frac{1}{\xi - \zeta} d\zeta d\xi.$$

If the last integral is broken into two parts, the results of the theorem are readily obtained from a trivial integration in the complex plane.

It should be noted that the conditions imposed by the present theorem on $C$ and $C'$ could be weakened without impairing the results (6) and (7). It is sufficient for (6) to require that $C$ and $C'$ be deformable within $\mathfrak{R}$ to two new curves which lie one inside the other. Likewise, (7) holds if $C$ and $C'$ are deformable within $\mathfrak{R}$ to two new curves which lie exterior to each other. All this is possible by virtue of Theorem 4.

The next theorem announces the most important property of the transformation $P$.

THEOREM 6. *The transformation $P$ of Theorem 4 is a projection, that is, $P^2=P$. Furthermore, it reduces $T$, that is, $PT=TP$.*

Let $C'$ be a curve in $\mathfrak{R}$ such that the curve $C$ defining $P$ lies entirely in the interior of $C'$ and such that $C$ and $C'$ may be deformed into each other within $\mathfrak{R}$. Let $P'$ be the transformation (4) associated with $C'$. By Theorem 4, $P=P'$. By Theorem 5, $PP'=P$. Hence $P^2=P$.

That $PT=TP$ has already been stated in Theorem 4. The significance of the role of $P$ relative to $T$ now becomes clear. The space $\mathfrak{B}$ may be thought of as the direct sum of two spaces $\mathfrak{M}$ and $\mathfrak{N}$, $\mathfrak{B}=\mathfrak{M}\dot{+}\mathfrak{N}$, with $\mathfrak{M}$ the totality of elements $f\in\mathfrak{B}$ for which $Pf=f$, and with $\mathfrak{N}$ the totality of elements $g\in\mathfrak{B}$ for which $Pg=0$. Since $PTf=TPf=Tf$, $T\mathfrak{M}\subset\mathfrak{M}$. In the same manner $T\mathfrak{N}\subset\mathfrak{N}$. Thus the study of $T$ in $\mathfrak{B}$ reduces to the study of $T$ in the two spaces $\mathfrak{M}$ and $\mathfrak{N}$.

**A special case.** A special case of considerable interest throws light on the character of the elements $f$ for which $Pf=f$ (or $Pf=0$). This character is completely revealed by the behavior of the sequence of iterates $T^nf$.

THEOREM 7. *Let the unit circle $C$ with center at the origin of the complex plane lie entirely in the resolvent set $\mathfrak{R}$ of $T$. Then the projection $P$ defined by* (4) *satisfies the relation*

$$P = \lim_{n\to\infty} (I - T^n)^{-1}. \tag{8}$$

*Furthermore,* $Pf=f$ *if and only if* $\lim_{n\to\infty}\|T^nf\|=0$. *And if* $Pf=0$, $f\neq 0$, *then* $\lim_{n\to\infty}\|T^nf\|=\infty$.

*If* $T^{-1}$ *exists, the projection associated by* (4) *to the transformation* $T^{-1}$ *and for the unit circle* $C$ *is* $I-P$ *where* $P$ *is as above. The elements* $f$ *for which* $Pf=0$ *are precisely those for which* $\lim_{n\to\infty}\|T^{-n}f\|=0$.

*In addition, if* $f$ $(\neq 0)$ *is such that* $\lim_{n\to\infty}\|T^nf\|=0$, *then* $\lim_{n\to\infty}\|T^{-n}f\|=\infty$. *If* $f$ $(\neq 0)$ *is such that* $\lim_{n\to\infty}\|T^{-n}f\|=0$ *then* $\lim_{n\to\infty}\|T^nf\|=\infty$.

The integral (4) in the special case indicated by the present theorem may be evaluated in the following manner: Let $\alpha=\exp(2\pi i/n)$. Consider as an approximation to the required integral the finite sum

$$\frac{1}{2\pi i}\sum_{j=0}^{n-1}(\alpha^jI-T)^{-1}(\alpha^{j+1}-\alpha^j).$$

Replace this sum by its equivalents

$$\frac{1}{2\pi i}\sum_{j=0}^{n-1}(I-\alpha^{-j}T)^{-1}(\alpha-1)=\frac{n}{2\pi i}(\alpha-1)(I-T^n)^{-1}.$$

Since $\lim_{n\to\infty}n(\alpha-1)=2\pi i$, equation (8) has been derived.

The relation

$$(I-T^n)^{-1}=I+(I-T^n)^{-1}T^n \tag{9}$$

may be verified immediately by multiplication with $(I-T^n)$. It implies that if $\lim_{n\to\infty}\|T^nf\|=0$, then since $|(I-T^n)^{-1}|$ is bounded (by (8)), it follows that $\lim_{n\to\infty}(I-T^n)^{-1}T^nf=0$. Applying (9) once more, this means $\lim_{n\to\infty}(I-T^n)^{-1}f=f$. In resumé, if $\lim_{n\to\infty}\|T^nf\|=0$, then $Pf=f$.

Now, assume that $Pf=f$. Write $(I-T^n)^{-1}=Q_n$ and let $n_0$ be so chosen that $|P-Q_n|<\epsilon<1/2$ for $n>n_0$. Let $Q_nf=g_n$. Then since $PT^n=T^nP$, $PT^nf=T^nf$, and (9) may be written

$$g_n-f=T^nf+(Q_n-P)T^nf.$$

Upon taking norms, this yields $\epsilon\|f\|\geqq\|T^nf\|-\epsilon\|T^nf\|$. Finally, $\|T^nf\|\leqq\epsilon/(1-\epsilon)\|f\|$ for $n>n_0$. This shows that $\lim_{n\to\infty}\|T^nf\|=0$.

If $Pf=0$, (9) yields $Q_nf=f+Q_nT^nf=f+(P-Q_n)T^nf+PT^nf$, and hence since $PT^nf=T^nPf=0$, $\|Q_nf\|+\epsilon\|T^nf\|\geqq\|f\|$. This gives $\|T^nf\|\geqq(1-\epsilon)/\epsilon\|f\|$ which means either $f=0$ or $\lim_{n\to\infty}\|T^nf\|=\infty$.

If $T^{-1}$ exists, the complex unit circle is in the resolvent set of $T^{-1}$ and (9) shows that $\lim_{n\to\infty}(I-T^{-n})^{-1}=I-P$. Proof of the remaining statements of the theorem may now be based on what has preceded.

**Some elementary properties of the fundamental projections.** The relationship between the spectrum of $T$ and the projection $P$ is given in the next theorem.

THEOREM 8. *The projection $P$ defined by the curve $C$ as in Theorem 4 has the following properties*:

(a) *If $\lambda$ is in the point spectrum of $T$, $Tf=\lambda f$ with $f\neq 0$, then $Pf=f$ or $Pf=0$ according as $\lambda$ lies within $C$ or without $C$.*

(b) *If $\lambda$ is in the continous spectrum and $\{f_n\}$ is a sequence of elements in $\mathfrak{B}$ with the properties $\|f_n\|=1$, $n=1, 2, \cdots$, $\lim_{n\to\infty}\|Tf_n-\lambda f_n\|=0$, then $\lim_{n\to\infty}\|Pf_n-f_n\|=0$ if $\lambda$ lies within $C$; $\lim_{n\to\infty}\|P_nf\|=0$ if $\lambda$ lies without $C$.*

(c) *The equation $P=0$ is valid if and only if every point in the interior of $C$ lies in the resolvent set of $T$.*

(d) *The equation $P=I$ is valid if and only if every point in the exterior of $C$ is in the resolvent set of $T$.*

The proof of (a) is immediate; that of (b) rests on the existence of elements $f$ for which $Tf=\lambda f$ approximately. In (c) if the interior of $C$ is in $\mathfrak{R}$, $C$ may be shrunk to a point, hence $P=0$ by Theorem 4. If $P=0$, application of (a) and (b) excludes the point and continuous spectrum from the interior of $C$. The residual spectrum is treated by reverting to the space ($\mathfrak{B}$). For (d), if $P=I$, clearly, the exterior of $C$ lies in $\mathfrak{R}$. In the converse case, assume $C$ to be a circle with center in $\mathfrak{R}$ and map its exterior upon the interior of the unit circle, then apply Theorem 7.

A by-product of this theorem is that every transformation has at least one point in its spectrum, a fact that has already been noted(7).

## III. THE HOMOMORPHISM OF SPECTRAL SETS AND PROJECTIONS

**Spectral sets.** Let $C$ be a simple closed rectifiable curve lying entirely in the resolvent set of $T$. The curve $C$ determines two sets of points, its interior $\mathbf{C}^i$, and its exterior $\mathbf{C}^e$. The totality of sets formed from the sets $\mathbf{C}^i$ and $\mathbf{C}^e$ (for all possible curves $C$) by the operations of complementation, finite set addition, and finite set intersection forms a Boolean algebra of sets. Any set in this algebra of sets will be called a *spectral set* of $T$.

With each spectral set $\boldsymbol{M}$ will be associated a projection $P_{\boldsymbol{M}}$ which reduces $T$. The method of assigning $P_{\boldsymbol{M}}$ to $\boldsymbol{M}$ will be the following: To the interior $\mathbf{C}^i$ of $C$ will be assigned the projection $P$ of Theorem 4. To the exterior $\mathbf{C}^e$ of $C$ will be assigned the projection $I-P$. To the set complements of $\mathbf{C}^i$ and $\mathbf{C}^e$ will be assigned the projections $I-P$ and $P$, respectively. To the intersection of the sets $\mathbf{C}_1^{u_1}, \mathbf{C}_2^{u_2}, \cdots, \mathbf{C}_n^{u_n}$, where the $u_i$ are symbols to denote an "$e$," an "$i$," or the complement of an "$e$" or "$i$," will be assigned the product of the corresponding projections. As the projections are commutative, the order of the terms in the product does not play any role. To a sum $\mathbf{C}_1^{u_1}+\mathbf{C}_2^{u_2}+\cdots+\mathbf{C}_n^{u_n}$ will be assigned the "starred sum" of the corresponding projections. The starred sum of the commutative projections $P_1, \cdots, P_n$ is defined by $P_1 \dotplus \cdots \dotplus P_n = I-(I-P_1)\cdots(I-P_n)$.

(7) Taylor, loc. cit.

Proceeding in this fashion, it is seen that if a spectral set $M$ is defined in a specific way by the formation of sums and intersections of the elemental sets $C_j^{u_i}$, a method is specified for forming the projection associated with $M$. That procedure consists in replacing in the definition of $M$ the sets $C_j^{u_i}$ by their projections, the operation of set intersection by that of projection multiplication, and the operation of set addition by that of projection starred addition. Finally if the projection $P$ is associated with the set $M$, the projection $I-P$ will be associated with the set complementary to $M$. Thus corresponding to a specified construction for a spectral set $M$, there is precisely one projection $P$ defined.

Although the projection associated with $M$ seems to depend on the particular manner in which $M$ is constructed, this is not the case. The proof, which will not be set down, proceeds along these lines. In any two methods of defining $M$, one replaces the auxiliary curves $C_i$ by simple polygons $D_i$ which are obtained from the $C_i$ by deformation. These $D_i$ generate in the plane a finite number of polygonal regions $Q_j$. The set $M$ by either method of definition consists (up to deformations) of certain of these regions. Finally, any region which lies in one deformation of $M$ but not in the other lies in $\mathfrak{R}$. Henceforth the projection associated with $M$ will be denoted by $P_M$.

THEOREM 9. *The mapping $M \rightarrow P_M$ of the set algebra of spectral sets on the associated class of projections is a homomorphism. Under the homomorphism $P_M=0$ if and only if $M$ lies entirely in the resolvent set of $T$.*

The establishment of the homomorphism is virtually accomplished by the discussion just preceding. Let $M_1$ and $M_2$ be two spectral sets. Then one method of defining the projection associated to $M_1+M_2$ (or $M_1 \cdot M_2$) is $P_{M_1+M_2}=P_{M_1} \dotplus P_{M_2}$ ($P_{M_1 \cdot M_2}=P_{M_1} \cdot P_{M_2}$). But this is precisely the homomorphism property. It is to be noted that since the projection associated with the whole plane is the identity, the projection $P_M$ associated with the set complement $\overline{M}$ of $M$ satisfies $P_{\overline{M}}=I-P_M$. That the sets $M$ for which $P_M=0$ are precisely those which contain no points of $\mathfrak{S}$ is a consequence of the uniqueness argument just preceding this theorem.

**Spectral components.** To obtain an idea of the resolving power of spectral sets, that is to say, of their ability to separate various parts of the spectrum, a few facts concerning sets may be recalled. Since the spectrum of a bounded transformation is closed and compact (or bounded) and since furthermore any closed compact set may serve as the spectrum of some bounded transformation, it is the general closed compact set which must be discussed. If $M$ is such a set and $a$ is any point in $M$, the set sum of all continua in $M$ containing $a$ is a continuum called a component of $M$([8]). Thus $M$ may be expressed as the sum of a finite or infinite number of distinct components. Since a component is a closed compact set, each two components lie at a positive distance

([8]) Hausdorff, *Mengenlehre*, p. 152.

from each other. Furthermore, if a simple closed curve has no point in common with a given component, that component lies entirely inside or outside the given curve.

If $K_1$ and $K_2$ are two distinct components (assuming that $M$ consists of more than one component), then there exist closed sets $M_1$ and $M_2$ such that $M_1 \cdot M_2 = 0$, $M = M_1 + M_2$, $M_1 \supset K_1$, $M_2 \supset K_2$(9). Finally, for the sets $M$, $M_1$, and $M_2$ there exists a finite set of polygonal domains having the properties: No point of $M$ is on the boundary of any domain; within any domain there is at least one point of $M_1$; all the points of $M_2$ lie without these domains(10). If $L$ represents a domain containing one point of $K_1$, $L$ contains $K_1$ in its entirety. Thus $L$ is a spectral set separating $K_1$ from $K_2$.

**The non-extensibility of the homomorphism.** A natural question to propose is whether the class of spectral sets $M$ and the projections $P_M$ can be enlarged in such a way that an extended homomorphism reigns between the two extended classes. Specifically, can this extension be carried out in such a fashion that the new homomorphism is valid for denumerable sums and intersections? Preliminary considerations show that it will be wise to forego the insistence that a set $M$ correspond to a projection $P_M$ and substitute the requirement that $M$ correspond to a pair of closed linear manifolds $\{\mathfrak{M}_M, \mathfrak{N}_M\}$ having the properties: $\mathfrak{M}_M$ and $\mathfrak{N}_M$ have no common element except 0; together they span $\mathfrak{B}$. The notion of a projection $P_M$ is a special case of this type where $\mathfrak{M}_M$ and $\mathfrak{N}_M$ are disjoint(11). $\mathfrak{M}_M$ is the set of elements $\{P_M f\}$, $f \in \mathfrak{B}$; $\mathfrak{N}_M$ is the set of elements $\{f - P_M f\}$, $f \in \mathfrak{B}$. The general type of manifold pair described above seems to play the leading role in the theory of rotations—inter alia—in reflexive spaces(12).

Further considerations suggest that the heretofore unqualified nature of the underlying space be restricted suitably so that it should have the "correct" properties relative to infinite intersections and sums of closed linear manifolds. Reflexive spaces seem to possess precisely the requisite properties(13).

---

(9) R. L. Moore, *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, p. 21, Theorem 35.

(10) Kerékjártó, *Vorlesungen über Topologie*, p. 31.

(11) $\mathfrak{M}_M$ and $\mathfrak{N}_M$ are disjoint if the elements $f+g$, $f \in \mathfrak{M}_M$, $g \in \mathfrak{N}_M$, are not only dense in $\mathfrak{B}$ but actually fill $\mathfrak{B}$.

(12) See the author's *The integral representation of weakly almost-periodic transformations in reflexive vector spaces*, these Transactions, vol. 49 (1941), pp. 18–40.

(13) If $(\mathfrak{B})$ represents the space adjoint to $\mathfrak{B}$, then the reflexive property is defined by $((\mathfrak{B})) = \mathfrak{B}$. Thus, for example, the spaces $L_p$ and $l_p$, $p > 1$, are reflexive. An important property alluded to is the following: If $\{\mathfrak{M}_n\}$ is a monotone decreasing sequence of closed linear manifolds and $\mathfrak{M}_n^\perp$ represents the largest manifold in $(\mathfrak{B})$ which is orthogonal to $\mathfrak{M}_n$, then $(\prod_{n=1}^{\infty} \mathfrak{M}_n)^\perp = \sum_{n=1}^{\infty} \mathfrak{M}_n^\perp$ where $\sum_{n=1}^{\infty} \mathfrak{M}_n^\perp$ indicates the smallest closed linear manifold containing each $\mathfrak{M}_n^\perp$.

*Proof.* Since $\mathfrak{M}_n^\perp \perp \mathfrak{M}_n$, then $\mathfrak{M}_n^\perp \perp \mathfrak{M} = \prod_{n=1}^{\infty} \mathfrak{M}_n$ or $\mathfrak{M}^\perp \supset (\mathfrak{N}) = \sum_{n=1}^{\infty} \mathfrak{M}_n^\perp$. Suppose $F \in \mathfrak{M}^\perp$, $F \notin (\mathfrak{N})$. Then since $\mathfrak{B}$ is reflexive, there exists an $f \in \mathfrak{B}$ such that $Ff = 1$ and $f \perp (\mathfrak{N})$. In particular, $f \perp \mathfrak{M}_n^\perp$, hence $f \in \mathfrak{M}_n$, $n = 1, 2, \cdots$. Thus $f \in \mathfrak{M}$ and $Ff = 0$. This contradiction proves that $\mathfrak{M}^\perp = (\mathfrak{N})$.

An example will be given of a simple transformation in Hilbert space $\mathfrak{H}$ (which is reflexive) to indicate that the desired extension of the homomorphism is impossible. Let $A$ be the linear transformation defined by the matrix $\|a_{ij}\|$, $i, j=0, 1, 2, \cdots$, with $a_{0j}=0, j=0, 1, 2, \cdots$; $a_{i0}=i\cdot 2^{-i}, i=1, 2, \cdots$; and $a_{ij}=\delta_{ij}\cdot 2^{-i}$, $i, j=1, 2, \cdots$. Since $\sum_{i,j=0}^{\infty}|a_{ij}|^2<\infty$, the transformation is of finite norm, hence certainly bounded. The spectrum of $A$ is found by examination of the equation

$$A(x_0, x_1, \cdots) - \lambda(x_0, x_1, \cdots) = (y_0, y_1, \cdots) \tag{10}$$

where $\sum_{n=0}^{\infty}|x_n|^2<\infty$, $\sum_{n=0}^{\infty}|y_n|^2<\infty$. It is found that if $\lambda$ does not assume values in the set $M=\{0, 1/2, 1/4, \cdots, 1/2^n, \cdots\}$, then (10) has a unique solution $(x_0, x_1, \cdots)$ for all $(y_0, y_1, \cdots)$. Thus the values in the set complementary to $M$ belong to the resolvent set of $A$. Also, $\lambda=1/2^n$, $n=1, 2, \cdots$ is found to be a characteristic value for a single vector $(x_0, x_1, \cdots)$ where $x_n=1$, otherwise $x_j=0$. Finally, the value $\lambda=0$ is in the residual spectrum since $Af=0$ implies $f=0$ and since the range of the transformation $A$ is not dense in space.

For the adjoint transformation $\overline{A}$, it is found that the value $\lambda=1/2^n$, $n=1, 2, \cdots$, is in the point spectrum and that the one characteristic vector for that value is the vector $(x_0, x_1, \cdots)$ where $x_0=1$, $x_n=1/n$, otherwise $x_j=0$. The value $\lambda=0$ is in the point spectrum of $\overline{A}$ and $\overline{A}(1, 0, 0, \cdots)=(0, 0, \cdots)$.

If $C_n$ represents a curve containing in its interior all points of $M$ save $1/2, 1/4, \cdots, 1/2^n$, and if $\overline{P}_n$ represents the projection associated with $C_n$ and $\overline{A}$, then the manifold $\{\overline{P}_nf\}, f\in\mathfrak{H}$, includes the manifold $\overline{\mathfrak{M}}_n$ of all elements of the form $(x_0, 0, 0, \cdots, 0, x_{n+1}, x_{n+2}, \cdots)$ by Theorem 8(a). Similarly, the manifold $\{f-\overline{P}_nf\}, f\in\mathfrak{H}$, includes the manifold $\overline{\mathfrak{N}}_n$ spanned by the $n$ elements $(1, 1/2, 0, 0, \cdots), (1, 0, 1/3, 0, \cdots)$ and $(1, 0, \cdots, 0, 1/n, 0, \cdots)$. As $\overline{\mathfrak{M}}_n$ and $\overline{\mathfrak{N}}_n$ span $\mathfrak{H}$, the manifold $\{\overline{P}_nf\}, f\in\mathfrak{H}$ is precisely $\overline{\mathfrak{M}}_n$, and the manifold $\{f-\overline{P}_nf\}, f\in\mathfrak{H}$, is precisely $\overline{\mathfrak{N}}_n$.

If one writes $\mathfrak{M}_n=\overline{\mathfrak{N}}_n^{\perp}$ and $\mathfrak{N}_n=\overline{\mathfrak{M}}_n^{\perp}$, one sees that the projection $P_n$ associated with the curve $C_n$ and with $A$ determines the manifolds $\mathfrak{M}_n$ and $\mathfrak{N}_n$. Now $\prod_{n=1}^{\infty}\mathfrak{M}_n=(\sum_{n=1}^{\infty}\overline{\mathfrak{N}}_n)^{\perp}=\mathfrak{H}^{\perp}=0$. Also $\sum_{n=1}^{\infty}\mathfrak{N}_n=(\prod_{n=1}^{\infty}\overline{\mathfrak{M}}_n)^{\perp}=\{(x_0, 0, 0, \cdots)\}^{\perp}\neq\mathfrak{H}$. This example shows that a homomorphism for denumerable intersections and sums cannot be expected. Also to be noted is that if one attempts to obtain the linear manifold which corresponds to the component of $M$ consisting of the point $\lambda=0$, one obtains the zero manifold, although the transformation has no singularities in this manifold!

As is fairly apparent, the contour integral used as above does not yield all possible projections which reduce $T$. This may be inferred easily from examples. In the first place, there is the possibility of multiplicity in the spectrum. Or even if the spectrum is simple, the integral may not "split" any

spectral component. Here one may refer for an example to the general rotation in a reflexive space.

There is a way, nevertheless, of obtaining readily a somewhat finer resolution into reducing manifolds than that exhibited above. If $f$ is any element in $\mathfrak{B}$, let $\mathfrak{M}_f$ represent the closed linear manifold spanned by the elements $Uf$ where $U$ is any function of $T$ (one may let $U$ represent any rational function of $T$). Then $T\mathfrak{M}_f \subset \mathfrak{M}_f$ and one may proceed as above replacing $\mathfrak{B}$ in the previous argument by $\mathfrak{M}_f$. It is to be noted that this procedure does not supply one with a manifold complementary to $\mathfrak{M}_f$. An obvious approach to problems of the type here envisaged is through a detailed study of the structure of the class of manifolds $\mathfrak{M}_f$ where $f$ ranges over $\mathfrak{B}$. This has not yet proved successful.

BARNARD COLLEGE, COLUMBIA UNIVERSITY,
NEW YORK, N. Y.

# ON THE OSCILLATION OF THE DERIVATIVES OF A PERIODIC FUNCTION

BY

GEORGE PÓLYA AND NORBERT WIENER

1. *Let $f(x)$ be a real valued periodic function of period $2\pi$ defined for all real values of $x$ and possessing derivatives of all orders. Let $N_k$ denote the number of changes of sign of $f^{(k)}(x)$ in a period. We consider the order of magnitude of $N_k$ as $k\to\infty$.*

(I) *If $N_k=O(1)$, $f(x)$ is a trigonometric polynomial.*

(II) *If $N_k=O(k^\delta)$ where $\delta$ is fixed, $0<\delta<1/2$, $f(x)$ is an entire function of finite order not exceeding $(1-\delta)/(1-2\delta)$.*

(III) *If $N_k=o(k^{1/2})$, $f(x)$ is an entire function.*

We prove this theorem by consideration of the Fourier series of $f(x)$

$$f(x) = \sum c_n e^{inx}, \tag{1}$$

$c_{-n}=\bar{c}_n$ $(n=0, 1, 2, \cdots)$. Here, as in what follows, the sign $\sum$ without explicitly stated limits means a summation from $-\infty$ to $\infty$. Under the present conditions, the series (1) is absolutely and uniformly convergent for real $x$, and so are the Fourier series of $f'(x), f''(x), \cdots$, obtained from (1) by term by term differentiation. If we focus our attention on the Fourier series, we may express the general trend of our theorem by saying that *a small amount of oscillation in the higher derivatives implies a rapid decrease in the coefficients,* this decrease being so extreme in case (I) that all coefficients from a certain point onward vanish.

The theorem we have to prove and a few analogous facts([1]) point towards a general principle which cannot yet be stated in precise terms but which is not entirely unsuitably expressed by saying that *a small amount of oscillation in the higher derivatives indicates a great amount of simplicity in the analytic nature of the function.*

An analogous theorem may be formulated for almost periodic functions. As in other theorems of this kind, the number of changes of sign in a period is replaced by the density of these changes over the infinite line and a trigonometric polynomial is replaced by an entire function of exponential type. The extension of case (I) of our theorem offers the least difficulty.

---

Presented to the Society, May 2, 1941; received by the editors August 8, 1941.

([1]) See S. Bernstein, *Leçons sur les Propriétés Extrémales*, Paris, 1926, pp. 190–197 and Communications de la Société Mathématique de Kharkow, (4), vol. 2 (1928), pp. 1–11; R. P. Boas and G. Pólya, Proceedings of the National Academy of Sciences, vol. 27 (1941), pp. 323–325.

2. We start with a few preliminary remarks on changes of sign. We consider first a real-valued function $f(x)$ which is defined in an interval $a \leqq x \leqq b$. We say that this function has $N$ changes of sign in this interval if it is possible to find $N+1$, and no more, abscissae $x_0, x_1, \cdots, x_N$ in the interval such that

$$x_0 < x_1 < \cdots < x_{N-1} < x_N, \tag{2}$$

$$f(x_{\nu-1})f(x_\nu) < 0, \qquad \nu = 1, 2, 3, \cdots, N. \tag{3}$$

*If a function has N changes of sign in an interval, its derivative has there at least $N-1$ changes of sign.* This variant of Rolle's theorem is easily proved by considering $\xi_\nu$, such that

$$f(x_\nu) - f(x_{\nu-1}) = (x_\nu - x_{\nu-1})f'(\xi_\nu), \qquad x_{\nu-1} < \xi_\nu < x_\nu,$$

and observing that $f(x_\nu)f'(\xi_\nu) > 0$ and that therefore

$$f'(\xi_{\nu-1})f'(\xi_\nu) < 0, \qquad \nu = 2, 3, \cdots, N.$$

Applying this to the function $e^{ax}f(x)$ (where $a$ is a real constant) and its derivative $[e^{ax}f(x)]' = e^{ax}[af(x) + f'(x)]$, we see that *the number of changes of sign of*

$$(a + D)f(x)$$

(where $D$ is the symbol of differentiation) *is not inferior to $N-1$, $N$ being the number of changes of sign of $f(x)$.*

Now let $f(x)$ be periodic with the period $2\pi$. We say that the number of changes of sign of $f(x)$ *in a period* is $N$, if it is possible to find just $N+1$, and no more, abscissae $x_0, x_1, \cdots, x_N$ such that

$$x_N = x_0 + 2\pi,$$

and (2), (3) hold. Observe that $f(x_N) = f(x_0)$ and that, therefore, $N$ is necessarily *even*. Hence it follows that the number of changes of sign of $(a+D)f(x)$ in a period is not inferior to that of $f(x)$. We defined $N_k$ in our initial statement; now we see that

$$N_0 \leqq N_1 \leqq N_2 \leqq \cdots \leqq N_{k-1} \leqq N_k \leqq \cdots. \tag{4}$$

Observing that

$$(a + D)(a - D) \sum \frac{c_n e^{inx}}{a^2 + n^2} = \sum c_n e^{inx},$$

we obtain:

Lemma I. *The number of changes of sign of the function* (1) *in a period is not inferior to that of*

$$\sum \frac{c_n e^{inx}}{a^2 + n^2}.$$

3. The series (1) represents a trigonometric polynomial of order $m$ if $c_n=0$ for $n=m+1, m+2, \cdots$. If $f(x)$ is a trigonometric polynomial of order $m$, it cannot have more than $2m$ roots in a period; this is well known. Observe that for large $k$, $f^{(k)}(x)$ has actually $2m$ changes of sign, because as $k\to\infty$, $(im)^{-k}f^{(k)}(x)$ approaches the first or the second of the two expressions

$$-c_{-m}e^{-imx}+c_me^{imx},\qquad c_{-m}e^{-imx}+c_me^{imx}, \tag{5}$$

according as $k$ is odd or even. The second of these expressions is of the form $2|c_m|\cos(mx-\gamma)$, with a certain real $\gamma$ and the first is of the same form except for a factor $i$.

The case (I) of our theorem characterizes the trigonometric polynomials and can be stated as follows: *A real-valued periodic function $f(x)$ possessing derivatives of all orders is a trigonometric polydomial if and only if the number of changes of sign of $f^{(k)}(x)$ remains bounded for $k\to\infty$.*

In order to prove this we consider (1). We have to show that some $f^{(k)}(x)$ have an arbitrarily great number of changes of sign if there are $c_n\neq 0$ with arbitrarily large subscripts $n$. More precisely we shall show this:

*If $m>0$ and $c_m\neq 0$, then all derivatives of* (1), *from a certain stage onward, have not less than $2m$ changes of sign.*

In fact, by repeated application of Lemma I, we ascertain that

$$f^{(k)}(x)=i^k\sum n^kc_ne^{inx} \tag{6}$$

does not have fewer changes of sign than

$$i^k\sum\left(\frac{2mn}{m^2+n^2}\right)^kc_ne^{inx}. \tag{7}$$

But since it is given that $c_m\neq 0$ and that $\sum c_n$ is absolutely convergent, we have from a certain $k$ onward

$$|c_m|>\left(\sum_{n=1}^{m-1}+\sum_{n=m+1}^{\infty}\right)\left(\frac{2mn}{m^2+n^2}\right)^k|c_n|. \tag{8}$$

Indeed we have for $n>0$, $n\neq m$, $0<2mn<m^2+n^2$, and therefore, each term tends to 0 on the right-hand side of (8) for $k\to\infty$.

But if (8) holds for a certain even $k$, the sum in (7) has the same sign as the second expression (5) in all those real points $x$ in which this latter reaches $2|c_m|$, the maximum of its absolute value. This maximum is reached with alternating signs, in equidistant points, the distance of two consecutive points being $\pi/m$. Therefore (7) has not less than $2m$ changes of sign. We have proved this for even $k$ but the same is true and the proof is nearly the same for odd $k$. Then, by Lemma I, (6) has not less than $2m$ changes of sign, and case (I) of our theorem is proved.

4. We consider the case (III) of our theorem before case (II).

If the periodic function $f(x)$ is analytic along the whole real axis, it is analytic in a certain horizontal strip bisected by the real axis and the Fourier series (1), which is a Laurent series in

$$z = e^{ix},$$

converges in the interior of the strip. Hence, by examining

$$\limsup_{n\to\infty} |c_n|^{1/n} \tag{9}$$

we can distinguish the following three cases:

If $f(x)$ is not analytic along the whole real axis, (9) has the value 1.

If $f(x)$ is analytic in a certain horizontal strip of width $2h$ bisected by the real axis, but in no wider horizontal strip, (9) has the value $e^{-h}$.

If $f(x)$ is an entire function, (9) has the value 0.

In order to prove case (III) of our theorem, we have to show that in the first two cases $N_k = o(k^{1/2})$ is excluded. We prove the following statement.

*If there exists a positive number $\gamma$ such that*

$$\limsup_{n\to\infty} |c_n| e^{n\gamma} = \infty, \tag{10}$$

*then there exists a positive number g such that $f^{(k)}(x)$ has, for an infinity of values of k, not less than $(k/g)^{1/2}$ changes of sign.*

By the considerations of the foregoing section, $f^{(k)}(x)$ has certainly not less than $2m$ changes of sign if (8) holds. Using (10), we have to find an arbitrarily large $m$ and a corresponding $k$ such that (8) holds. We shall succeed in finding such an $m$ by applying the following known lemma(²).

LEMMA II. *We consider two infinite sequences $l_1, l_2, \cdots, l_n, \cdots$ and $s_1, s_2, \cdots, s_n, \cdots$, and suppose that*

$$l_n \geqq 0, \qquad n = 1, 2, 3, \cdots, \tag{11}$$

$$0 < s_1 < s_2 < s_3 < \cdots, \tag{12}$$

$$\lim_{n\to\infty} l_n = 0, \tag{13}$$

$$\limsup_{n\to\infty} l_n s_n = \infty. \tag{14}$$

*Then there exists an infinity of integers m such that*

$$l_m \geqq l_{m+\mu}, \qquad \mu = 1, 2, 3, \cdots,$$

$$l_m s_m \geqq l_{m-\mu} s_{m-\mu}, \qquad \mu = 1, 2, \cdots, m-1.$$

We put

(²) See G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, Berlin, 1925, vol. 1, pp. 18 and 173, Problem 109.

$$|c_n| = l_n, \quad e^{n\gamma} = s_n.$$

This choice satisfies (11), (12), (13), (14); in fact, (13) is satisfied because (1) is convergent, and (14) is satisfied because we have supposed (10). Thus we obtain an infinity of $m$ such that

$$|c_{m+\mu}| \leqq |c_m|, \qquad \mu = 1, 2, 3, \cdots,$$
$$|c_{m-\mu}| e^{(m-\mu)\gamma} \leqq |c_m| e^{m\gamma}, \qquad \mu = 1, 2, \cdots, m-1.$$

This we use to estimate the following sum. (Our ultimate aim is to prove (8).)

$$\tag{15} \begin{aligned} &\sum_{\mu=1}^{m-1}\left(\frac{2m(m-\mu)}{m^2+(m-\mu)^2}\right)^k \left|\frac{c_{m-\mu}}{c_m}\right| + \sum_{\mu=1}^{\infty}\left(\frac{2m(m+\mu)}{m^2+(m+\mu)^2}\right)^k \left|\frac{c_{m+\mu}}{c_m}\right| \\ &< \sum_{\mu=1}^{m-1}\left(\frac{2m(m-\mu)}{m^2+(m-\mu)^2}\right)^k e^{\gamma\mu} + \sum_{\mu=1}^{\infty}\left(\frac{2m(m+\mu)}{m^2+(m+\mu)^2}\right)^k \\ &= S_1 + S_2. \end{aligned}$$

We introduced the abbreviations

$$\tag{16} S_1 = \sum_{\mu=1}^{m-1} e^{\gamma\mu}\left(1 + \frac{\mu^2}{2m(m-\mu)}\right)^{-k},$$

$$\tag{17} S_2 = m\sum_{\mu=1}^{\infty}\left(\frac{2(1+\mu/m)}{1+(1+\mu/m)^2}\right)^k \frac{1}{m},$$

and we shall consider $S_1$ and $S_2$ in turn.

(1) Split the sum $S_1$ in two parts, $\mu$ being less than or equal to $m/2$ in the first part and greater than $m/2$ in the second. Using the fact that

$$(1+x)^{-1} < e^{-x/2}, \qquad 0 < x < 1,$$

we obtain

$$\begin{aligned} S_1 &< \sum_{1}^{m/2} e^{\gamma\mu - k\mu^2/4m^2} + \sum_{m/2}^{m-1} e^{\gamma\mu}(4/5)^k \\ &< \sum_{1}^{m/2} e^{-[(k/4m^2)-\gamma]\mu} + me^{\gamma m}(4/5)^k \\ &< e^{-(g-\gamma)}/(1-e^{-(g-\gamma)}) + me^{\gamma m}(4/5)^{4gm^2}. \end{aligned}$$

We put

$$\tag{18} k = 4gm^2$$

where $g$ is a positive integer, $g > \gamma$. We choose a fixed $g$ such that for sufficiently great $m$

$$\tag{19} S_1 < 1/2.$$

(2) The function $2x(1+x^2)^{-1}$ decreases for $x>1$. Therefore by (17)

$$S_2 < m\int_1^\infty \left(\frac{2x}{1+x^2}\right)^k dx \sim \frac{m}{2}\left(\frac{2\pi}{k}\right)^{1/2} = \frac{1}{2}\left(\frac{\pi}{2g}\right)^{1/2}.$$

We used a well known asymptotic evaluation of definite integrals([8]) and (18). If $g \geqq 2$, which we assume, we obtain for sufficiently great $m$

$$S_2 < 1/2. \tag{20}$$

But (15), (19), (20) show that (8) is true so that $f^{(k)}(x)$ has not fewer changes of sign than

$$2m = (k/g)^{1/2}.$$

5. We now proceed to the proof of case (II).

LEMMA III. *The Fourier series* (1) *represents an entire function of the finite order* $\lambda$, $\lambda > 1$, *if and only if*

$$\liminf_{n\to\infty} \frac{\log\log(1/|c_n|)}{\log n} = \frac{\lambda}{\lambda - 1}. \tag{21}$$

The proof consists of two parts. Both parts follow familiar lines; so we do not give all the details.

(1) Assume that $f(x)$ is entire and of order $\lambda$. Then for a fixed positive $\epsilon$ and for sufficiently great $|x|$,

$$|f(x)| < e^{|x|^{\lambda+\epsilon}}. \tag{22}$$

If we evaluate $c_n$ and shift the line of integration (periodicity and Cauchy's formula), we obtain as a result that if $r$ is any positive number,

$$\begin{aligned} c_n &= \frac{1}{2\pi}\int_{-ir-\pi}^{-ir+\pi} f(x)e^{-inx}dx \\ &= \frac{1}{2\pi} e^{-nr}\int_{-\pi}^{\pi} f(-ir+u)e^{-inu}du, \\ |c_n| &\leqq e^{(r+\pi)^{\lambda+\epsilon}-nr}. \end{aligned} \tag{23}$$

Here we use (22). We choose $r$, for given $n$, so that this right-hand side of (23) shall be a minimum. It follows by straight-forward calculation that (21) holds with "$\geqq$" instead of "$=$."

(2) Assume that

$$\liminf_{n\to\infty} \frac{\log\log(1/|c_n|)}{\log n} = \kappa > 0.$$

([8]) See, for example, G. Pólya and G. Szegö, loc. cit., vol. 1, pp. 78 and 244, Problem 201.

Therefore we have, for a given positive $\epsilon$ and all sufficiently great $n$

$$\left| c_n e^{inx} \right| < e^{-n^{\kappa-\epsilon}+|x|n}.$$

We choose $n$, for a given $x$, so that the right-hand side is a maximum. This maximum gives the right order of magnitude because the terms of (1) whose index surpasses a certain multiple of the index of the maximum term, yield a negligible contribution. We find that the order $\lambda$ of $f(x)$ satisfies the inequality

$$\lambda \leqq \frac{\kappa}{\kappa - 1}.$$

This gives (21) with "$\leqq$" instead of "$=$."

6. Now we are prepared to prove case (II) of our theorem. We have to show that *if the entire function $f(x)$ is of order $\lambda$, and $\epsilon > 0$ then, for an infinity of $k$,*

$$N_k > k^{(\lambda-1)/(2\lambda-1)-\epsilon}.$$

Put $\lambda/(\lambda-1)+\eta=\gamma$, $\eta$ being positive and small. By Lemma III, the fact we have to show can be stated as follows:

*If there exists a positive number $\gamma$, $\gamma > 1$, such that*

$$\limsup_{n\to\infty} \left| c_n \right| e^{n^\gamma} = \infty,$$

*then there exists a positive $g$ such that $f^{(k)}(x)$ has, for an infinity of $k$, more than $(k/g)^{1/(\gamma+1)}$ changes of sign.*

We apply Lemma II, whose conditions are satisfied by

$$l_n = \left| c_n \right|, \qquad s_n = e^{n^\gamma}.$$

We obtain the result that for an infinity of $m$,

$$\begin{aligned} \left| c_m \right| &\geqq \left| c_{m+\mu} \right|, & \mu &= 1\ 2, \cdots, \\ \left| c_m \right| e^{m^\gamma} &\geqq \left| c_{m-\mu} \right| e^{(m-\mu)^\gamma}, & \mu &= 1, 2, \cdots, m-1. \end{aligned}$$

Hence, using the fact that $\gamma > 1$, we obtain

$$\begin{aligned} &\sum_{\mu=1}^{m-1} \left( \frac{2m(m-\mu)}{m^2+(m-\mu)^2} \right)^k \left| \frac{c_{m-\mu}}{c_m} \right| + \sum_{\mu=1}^{\infty} \left( \frac{2m(m+\mu)}{m^2+(m+\mu)^2} \right)^k \left| \frac{c_{m+\mu}}{c_m} \right| \\ &\quad < \sum_{\mu=1}^{m-1} \left( \frac{2m(m-\mu)}{m^2+(m-\mu)^2} \right)^k e^{\gamma m^{\gamma-1}\mu} + \sum_{\mu=1}^{\infty} \left( \frac{2m(m+\mu)}{m^2+(m+\mu)^2} \right)^k \\ &\quad = S_1' + S_2. \end{aligned} \tag{24}$$

$S_2$ has the same meaning as before (see (17)), and

$$
\begin{aligned}
S_1' &= \sum_{\mu=1}^{m-1} e^{\gamma m^{\gamma-1}\mu}\left(1+\frac{\mu^2}{2m(m-\mu)}\right)^{-k} \\
&< \sum_{\mu=1}^{m/2} e^{\gamma m^{\gamma-1}\mu - k\mu^2/4m^2} + m e^{\gamma m^{\gamma}}(4/5)^{k} \qquad (25)\\
&= \sum_{\mu=1}^{m/2} e^{-m^{\gamma-1}(g\mu^2-\gamma\mu)} + m e^{\gamma m^{\gamma}}(4/5)^{4gm^{\gamma+1}}.
\end{aligned}
$$

We put

$$k = 4gm^{\gamma+1},$$

and we choose $g$ so that

$$\gamma + 1 \leqq g < \gamma + 2,$$

and so that $k$ is an integer. This choice assures that

$$S_1' \to 0, \qquad S_2 \to 0$$

for $m \to \infty$ (see (25) and the considerations preceding (20)). Therefore, by (24), (8) is true and $f^{(k)}(x)$ has not fewer changes of sign than

$$2m > 2[k/4(\gamma+2)]^{1/(\gamma+1)}.$$

Brown University,
Providence, R. I.
Massachusetts Institute of Technology,
Cambridge, Mass.

# STRUCTURE OF LINEAR SETS

BY

ERNST SNAPPER

## Introduction

In general only vector spaces whose scalar domains are fields have been investigated. In this paper we study vector spaces whose scalar domains are integral domains in which every ideal has a finite basis. It is shown that the theory of the linear subsets of such a vector space can be developed by using exactly the same technique which has always been applied to the ideals of an integral domain. This technique is established in the first two sections and is then used to develop the Noether decomposition theory of linear sets into "primary" linear sets.

## I. Basis theorem and ascending chain condition

We first define the notion of vector space:

DEFINITION 1.1. *An n-dimensional vector space consists of all the vectors with n components where the components form an integral domain in which every ideal has a finite basis.*

This integral domain is called the scalar domain of the vector space.

DEFINITION 1.2. *A linear set is a subset of an n-dimensional vector space closed under vector subtraction and multiplication by arbitrary scalars.*

An integral domain in which every ideal has a finite basis is clearly a one-dimensional vector space and its ideals are the linear sets. The ordinary theory of ideals will therefore become a part of this theory of linear sets.

**Notation.** The fixed $n$-dimensional vector space will be indicated by $V_n$ and its scalar domain by $R$. Otherwise the capital Latin letters will be used exclusively to indicate linear sets and the small Latin letters to indicate vectors. The ideals of $R$ will be indicated by German letters and the scalars of $R$ by Greek letters. The radical of an ideal $\mathfrak{a}$ will be indicated by $\mathfrak{a}'$. As is customary, $\mathfrak{a}'$ consists of all the scalars of which a power lies in $\mathfrak{a}$.

The following theorems, whose proofs are apparent from van der Waerden's *Moderne Algebra*, vol. 2 (we refer to this book as [1]), will be used frequently:

THEOREM 1.1 (The basis theorem). *Every linear set has a finite basis.*

Presented to the Society, February 22, 1941; received by the editors December 3, 1940, and, in revised form August 9, 1941.

THEOREM 1.2 (The ascending chain condition). *A chain of linear sets* $L_1 \subset L_2 \subset \cdots$ *in which* $L_i$ *is properly contained in* $L_{i+1}$ *is finite.*

THEOREM 1.3 (Maximum condition). *Every non-empty set of linear sets possesses a maximal set (that is, a linear set which is not contained in any other linear set of that set).*

THEOREM 1.4 (Induction by division). *If a property can be proved for every linear set* $L$, $V_n$ *included, under the assumption that the property can be proved for all the linear sets which properly contain* $L$, *the property holds for every linear set.*

## II. PRODUCTS AND QUOTIENTS

The technique mentioned above for treating the linear sets is based upon the following definitions:

DEFINITION 2.1. *The greatest common divisor or sum* $L = (L_1 + L_2)$ *of two linear sets* $L_1$ *and* $L_2$ *is the linear set* $L$, *generated by their logical sum.*

DEFINITION 2.2. *The least common multiple* $L = [L_1 \cap L_2]$ *of two linear sets* $L_1$ *and* $L_2$ *is their intersection.*

DEFINITION 2.3. *The product* $L = \mathfrak{a}L_1$ *of an ideal* $\mathfrak{a}$ *and a linear set* $L_1$ *is the linear set* $L$, *generated by the scalar products of scalars of* $\mathfrak{a}$ *and vectors of* $L_1$.

DEFINITION 2.4. *The quotient* $L = L_1/\mathfrak{a}$ *of a linear set* $L_1$ *and an ideal* $\mathfrak{a}$ *is the linear set* $L$, *consisting of the vectors whose scalar products with all the scalars of* $\mathfrak{a}$ *lie in* $L_1$.

DEFINITION 2.5. *The quotient* $\mathfrak{a} = L_1/L_2$ *of a linear set* $L_1$ *and a linear set* $L_2$ *is the ideal* $\mathfrak{a}$, *consisting of the scalars whose scalar products with all the vectors of* $L_2$ *lie in* $L_1$.

The theorems concerning these notions are the same as in the case of ideals and are proved in the same way. The most important ones are listed below (for notation see Part I):

THEOREM 2.1. *Multiplication is associative and is distributive with respect to addition*:

$$(\mathfrak{a}_1\mathfrak{a}_2)L = \mathfrak{a}_1(\mathfrak{a}_2L); \qquad (\mathfrak{a}_1 + \mathfrak{a}_2)L = (\mathfrak{a}_1L + \mathfrak{a}_2L); \qquad \mathfrak{a}(L_1 + L_2) = (\mathfrak{a}L_1 + \mathfrak{a}L_2).$$

THEOREM 2.2. *Division is distributive with respect to intersection*:

$$[L_1 \cap L_2 \cap \cdots \cap L_r]/M = [L_1/M \cap L_2/M \cap \cdots \cap L_r/M];$$
$$[L_1 \cap L_2 \cap \cdots \cap L_r]/\mathfrak{a} = [L_1/\mathfrak{a} \cap \cdots \cap L_r/\mathfrak{a}].$$

THEOREM 2.3. *The quotient of the sum is the intersection of the quotients*:

$$L/(\mathfrak{a}_1 + \mathfrak{a}_2) = [L/\mathfrak{a}_1 \cap L/\mathfrak{a}_2]; \qquad L/(M_1 + M_2) = [L/M_1 \cap L/M_2].$$

## III. Closure, essential ideal and radical

The notions discussed in this section play an important role in the structure theory of linear sets.

Definition 3.1. *The closure $\bar{L}$ of a linear set $L$ is the linear set consisting of all the vectors of which some scalar multiple, different from zero, lies in $L$.*

This definition is clearly consistent and we always have

$$L \subseteq \bar{L}, \qquad \overline{[L_1 \cap L_2]} = [\bar{L}_1 \cap \bar{L}_2].$$

Definition 3.2. *The linear set $L$ will be called closed if $\bar{L} = L$ and dense if $\bar{L} = V_n$.*

Example 3.1. If $R$ consists of all the rational integers, then in $V_2$ over $R$ the set $L_1$ generated by $(2, 2)$ is neither closed nor dense; $L_2 = (1, 1)$ is closed but not dense; $L_3$, the set generated by $(1, 0)$ and $(0, 2)$, is not closed but is dense.

Every closure of a linear set is a closed set and the nonzero ideals, considered as the linear subsets of a one-dimensional vector space, are dense sets. The only linear set which is both closed and dense is $V_n$.

The following theorem will be used later:

Theorem 3.1. *The intersection of two closed sets is closed. The intersection of two dense sets is dense. The irredundant intersection of a closed set and a dense set is neither closed nor dense. (An intersection is called irredundant if none of the intersection components is superfluous.)*

This theorem follows immediately from the fact that

$$\overline{[L_1 \cap L_2]} = [\bar{L}_1 \cap \bar{L}_2].$$

Definition 3.3. *The essential ideal $\mathfrak{E} = L/V_n$ of a linear set $L$ is the quotient ideal of the linear set and the whole space $V_n$. The radical $\mathfrak{E}'$ of $L$ is the radical of $\mathfrak{E}$.*

The essential ideal $\mathfrak{E}$ clearly consists of all the scalars which transform the whole space by means of scalar multiplication into the linear set. Therefore we have $L/\mathfrak{E} = V_n$. The essential ideal of an ideal is the ideal itself and the radical of an ideal is just its ordinary radical.

Theorem 3.2. *The following three statements are equivalent: $L$ is dense; $\mathfrak{E} \neq 0$; $L$ has maximal dimension $n$.*

**Proof.** If $L$ is dense, the ideals $L/(1, 0, \cdots, 0)$, $L/(0, 1, 0, \cdots, 0)$, $\cdots$, $L/(0, \cdots, 1)$ are nonzero ideals and therefore $\mathfrak{E} = L/V_n = [L/(1, 0, \cdots, 0) \cap \cdots \cap L/(0, \cdots, 1)]$ is a nonzero ideal. If $\mathfrak{E} \neq 0$, there exists a scalar $\lambda$ such that the $n$ vectors $(\lambda, 0, \cdots, 0) \cdots (0, \cdots, \lambda)$ lie in $L$ and therefore $L$

has maximal dimension $n$. If $L$ has maximal dimension $n$, there are $n$ independent vectors $f_1, \cdots, f_n$ in $L$. An arbitrary vector $v$ may then be written as $v=(\alpha_1/\beta_1)f_1+\cdots+(\alpha_n/\beta_n)f_n$ and therefore $\lambda v=\rho_1 f_1+\cdots+\rho_n f_n$ lies in $L$. Consequently, $L$ is dense. This proves the theorem.

We conclude this section by remarking that $L\subseteq L/\mathfrak{a}\subseteq\bar{L}$ and $\mathfrak{E}_L\subseteq L/M\subseteq R$, where $\mathfrak{a}$ has to be a nonzero ideal.

## IV. Primary linear sets and prime linear sets

Definition 4.1. *A linear set L will be called primary if* $\lambda v\equiv 0(L)$ *implies either* $v\equiv 0(L)$ *or* $\lambda\equiv 0(\mathfrak{E}')$.

Definition 4.2. *A linear set will be called prime if the linear set is primary and its* $\mathfrak{E}$ *is a prime ideal.*

It is clear that the ideals which are prime or primary according to these definitions are the ordinary prime and primary ideals.

Theorem 4.1. *If L is primary* $\mathfrak{E}$ *is primary and therefore* $\mathfrak{E}'$ *is prime.*

**Proof.** $\alpha\beta\equiv 0(\mathfrak{E})$ and $\beta\not\equiv 0(\mathfrak{E})$ imply $L/\alpha\beta=V_n$ and $L/\beta\neq V_n$. We can find therefore a vector $v$ such that $\alpha\beta v\equiv 0(L)$ and $\beta v\not\equiv 0(L)$ from which it follows that $\alpha\equiv 0(\mathfrak{E}')$, q.e.d.

The converse does not hold, since the set generated by the vector (2, 4) in the vector space whose scalar domain consists of the rational integers is not primary while its essential ideal is the zero ideal and therefore prime.

Theorem 4.2. *A primary linear set is either closed or dense.*

**Proof.** If $\mathfrak{E}\neq 0$, $L$ is dense. If $\mathfrak{E}=0$, $\mathfrak{E}'=0$, from which we conclude that if $\lambda v\equiv 0(L)$ and $v\not\equiv 0(L)$, $\lambda$ has to be zero. This means that $L$ is closed.

The converse does not hold since an ideal is a dense set but may not be primary. However, every closed set is a prime linear set since every closed linear set other than $V_n$ has a zero essential ideal.

The following theorems will demonstrate further that a primary linear set is the exact analogue of a primary ideal. These theorems are proved in exactly the same way as in the case of ideals and are used constantly in the structural theory of linear sets.

Theorem 4.3. *Let L be an arbitrary linear set,* $\mathfrak{E}$ *its essential ideal and* $\mathfrak{a}$ *an arbitrary ideal. Then L is primary and* $\mathfrak{a}$ *is equal to the radical* $\mathfrak{E}'$ *of the linear set if and only if the following three conditions hold:* (1) $\lambda v\equiv 0(L)$ *implies either* $v\equiv 0(L)$ *or* $\lambda\equiv 0(\mathfrak{a})$; (2) $\mathfrak{E}\subseteq\mathfrak{a}$; (3) $\lambda\equiv 0(\mathfrak{a})$ *implies* $\lambda^k\equiv 0(\mathfrak{E})$ *for some k.*

Theorem 4.4. *Let L be a primary linear set,* $\mathfrak{a}$ *an ideal and M a linear set. Then* $\mathfrak{a}M\equiv 0(L)$ *implies either* $M\equiv 0(L)$ *or* $\mathfrak{a}\equiv 0(\mathfrak{E}')$.

Corollary 4.1. *If L is primary and* $\mathfrak{a}\not\equiv 0(\mathfrak{E}')$, $L/\mathfrak{a}=L$.

THEOREM 4.5. *If* $L=[L_1\cap L_2\cap\cdots\cap L_r]$, *we have* $\mathfrak{E}=[\mathfrak{E}_1\cap\mathfrak{E}_2\cap\cdots\cap\mathfrak{E}_r]$ and $\mathfrak{E}'=[\mathfrak{E}_1'\cap\mathfrak{E}_2'\cap\cdots\cap\mathfrak{E}_r']$, *where* $\mathfrak{E}_i$ *and* $\mathfrak{E}_i'$ *are the essential ideal and radical of* $L_i$.

## V. THE NOETHER DECOMPOSITION THEORY

A linear set will of course be called reducible if it is the irredundant intersection of two other linear sets and irreducible if it is not reducible.

From Theorem 1.2, the following theorem is derived:

THEOREM 5.1. *Every linear set is the irredundant intersection of a finite number of irreducible linear sets.*

After each of the following theorems we have indicated how to derive the proofs of these theorems from the very similar proofs of the corresponding theorems on ideals, given in [1].

THEOREM 5.2. *Every irreducible linear set is primary.*

**Proof.** (See [1, pp. 31 and 32].) Replace the radical of the ideal by the radical of the linear set and the integral domain by the vector space.

An immediate corollary of Theorems 5.1 and 5.2 is:

THEOREM 5.3. *Every linear set is the irredundant intersection of a finite number of primary linear sets.*

THEOREM 5.4. *The intersection of a finite number of primary linear sets which all have the same radical is a primary linear set with that same radical.*

**Proof.** (See [1, pp. 32 and 33].) Use Theorem 4.3.

By replacing the intersection components which have the same radical by their intersection we prove:

THEOREM 5.5. *Every linear set is the irredundant intersection of a finite number of primary linear sets whose radicals are all different.*

A representation of a linear set as described in Theorem 5.5 is called "a representation by maximal primary components."

As a preparation for the first uniqueness theorem we need the following lemma:

LEMMA 5.1. *In two representations of a linear set by maximal primary components, each radical which is maximal among all the radicals of both the representations is present in both the representations.*

**Proof.** (See [1, pp. 35, 36].) Divide the components of both the representations by the essential ideal of a primary component whose radical is maximal among all the radicals of both the representations. Use Corollary 4.1.

From this lemma the following theorem is easily derived:

THEOREM 5.6. *An irredundant intersection of a finite number of primary linear sets, whose radicals are not all the same, is not primary.*

This theorem shows that in a representation of a linear set by maximal primary components no group of components has a primary intersection.

THEOREM 5.7 (The first uniqueness theorem). *In two representations of a linear set by maximal primary components, the number of components is the same and so are the radicals of the components of the two representations.*

**Proof.** (See [1, p. 36].) Single out, from the two representations, two linear sets $L_1$ and $L_2$ whose radicals $\mathfrak{E}_1'$ and $\mathfrak{E}_2'$ are the same and are maximal among all the radicals of both the representations. Then divide both the representations by the intersection of the two essential ideals of $L_1$ and $L_2$.

The uniquely determined radicals which occur in a representation of a linear set by maximal primary components are prime ideals (Theorem 4.1). They will be called the "associated prime ideals" of the linear set.

The following theorems are essential for the proof of the second uniqueness theorem:

THEOREM 5.8. *We have $M/\mathfrak{E}_L = M$ if and only if the essential ideal $\mathfrak{E}_L$ of the linear set L is not contained in any of the associated prime ideals of the linear set M.*

**Proof.** (See [1, p. 37].) Divide by $\mathfrak{E}_L$ and use the fact that $\mathfrak{E}_1 N_2 \subseteq [N_1 \cap N_2]$, where $N_1$ and $N_2$ are linear sets and $\mathfrak{E}_1$ is the essential ideal of $N_1$.

Another way of stating this theorem is:

THEOREM 5.9. *We have $M/\mathfrak{E}_L = M$, where $\mathfrak{E}_L$ is the essential ideal of L, if and only if no associated prime ideal of L is contained in an associated prime ideal of M.*

This theorem leads to:

DEFINITION 5.1. *The linear set L is relatively prime with respect to the linear set M if and only if $M/\mathfrak{E}_L = M$.*

The following definitions lead to the second uniqueness theorem:

DEFINITION 5.2. *An associated prime ideal of a linear set is said to be imbedded if it contains another associated prime ideal of that linear set.*

DEFINITION 5.3. *An intersection of maximal primary components, all belonging to the same representation of a linear set L by maximal primary components, is called a component linear set of L.*

DEFINITION 5.4. *Two component linear sets, which are taken from the same representation of a linear set L by maximal primary components and which have intersection L, are called conjugate component linear sets of L.*

If we call the conjugate of $L$, considered as a component linear set of itself, $V_n$, we have: Every component linear set of a linear set has a conjugate, which is not necessarily unique.

DEFINITION 5.5. *A component linear set of a linear set is said to be isolated if none of its associated prime ideals contains an associated prime ideal of its conjugate.*

THEOREM 5.10 (The second uniqueness theorem). *An isolated component linear set of a linear set is uniquely determined by its associated prime ideals.*

**Proof.** (See [1, p. 38].) Use Theorem 5.9. If in [1] the author divides by an ideal, here we have to divide by the essential ideal of the corresponding linear set.

## VI. TWO CONSEQUENCES OF THE STRUCTURE THEORY

It might be possible that the associated prime ideals of a linear set were always the same as the associated prime ideals of the essential ideal of that linear set. This would suggest that the representation of the linear set by maximal primary components were induced by the Noether decomposition of the essential ideal of the linear set. However, this conjecture will be disproved by the following counterexample:

EXAMPLE 6.1. Let $V_2$ have the polynomials in two variables $x$ and $y$ with integral coefficients as scalar domain. Let $L_1$ be the linear set consisting of all the vectors whose two components have a difference congruent to zero modulo $(x^2)$, and $L_2$ the linear set consisting of the vectors whose two components have a sum congruent to zero modulo $(x^2, y)$. Consider finally the linear set $L=[L_1\cap L_2]$. The following statements can easily be verified:

$L_1$ is a primary linear set. $\mathfrak{E}_1=(x^2)$ and $\mathfrak{E}_1'=(x)$.

$L_2$ is a primary linear set. $\mathfrak{E}_2=(x^2, y)$ and $\mathfrak{E}_2'=(x, y)$.

$L=[L_1\cap L_2]$ is a representation of $L$ by maximal primary components. The associated prime ideals of $L$ are $(x)$ and $(x, y)$. However, the essential ideal of $L$ is $(x^2)$ and its only associated prime ideal is $(x)$.

This example illustrates the general situation described in the following theorem:

THEOREM 6.1. *The associated prime ideals of the essential ideal of a linear set are among the associated prime ideals of that linear set.*

**Proof.** From Theorem 4.5 we see that the representation of a linear set by maximal primary components induces a decomposition of the essential ideal into primary ideals, whose radicals are all different and equal to the associated primes of the linear set. However, this decomposition may not be irredundant, as Example 6.1 shows, where this induced decomposition is $(x^2)=[(x^2)\cap(x^2, y)]$. By deleting the unnecessary components in the inter-

section, we get the Noether decomposition of the essential ideal and the remaining radicals are the associated primes of the essential ideal.

THEOREM 6.2. *Every linear set is the irredundant intersection of a closed set and a dense set. The closed part of the intersection is unique and is the closure of the set. The dense part of the intersection is not unique, even when the intersections are restricted to representations by maximal primary components. The dense part of the intersection is not present if and only if the linear set is closed, and the closed part of the intersection is not present if and only if the set is dense.*

**Proof.** Consider a representation of the linear set by maximal primary components. From Theorems 3.1 and 3.2 it follows that the intersection of the intersection components whose radicals differ from zero is a dense set and that the intersection component whose radical is the zero ideal is a closed set. This dense set and this closed set clearly satisfy the requirements of the theorem. That the closed part of the intersection is the closure of the linear set is proved as follows: Let $L=[C\cap D]$ where $D$ is dense and $C$ is closed. Then $\bar{L}=[\bar{C}\cap\bar{D}]$ (Part III) and $\bar{C}=C$ and $\bar{D}=V_n$. Therefore $\bar{L}=C$. The last part of the theorem follows immediately from Theorem 3.1. Finally, the following example shows that the dense part of the intersection is not unique, even if we restrict ourselves to representations by maximal primary components:

EXAMPLE 6.2. Let $V_2$ have the rational integers as scalar domain. Let $L$ be the linear set generated by the vector (2, 2). The following statements can easily be verified: The linear set generated by the vector (1, 1) is the closure $\bar{L}$ of $L$. $L$ is the intersection of $\bar{L}$ and the dense set generated by the vectors (2, 2) and (1, 3). $L$ is also the intersection of $\bar{L}$ and the dense set generated by the vectors (2, 2) and (2, 4). These are two representations of $L$ by maximal primary components and the two dense parts of the two intersections are different.

## REFERENCE

1. B. L. van der Waerden, *Moderne Algebra*, vol. 2, Berlin, 1940.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

# ON A GENERALIZATION OF THE PROBLEM OF QUASI-ANALYTICITY

BY
S. MANDELBROJT AND F. E. ULRICH

**Introduction.** The class of all functions analytic in the closed interval $a \leqq x \leqq b$, may be characterized in the following manner: It is the class of all functions $f(x)$ defined and infinitely differentiable (indicated hereafter as i.d. functions) in the interval such that to each function there corresponds a positive constant $k=k(f)$ with the property [5](1) that

$$|f^{(n)}(x)| < k^n n!, \qquad n \geqq 1;\ a \leqq x \leqq b.$$

Gevray(2), in studying the heat equation, introduced functions $\Phi(x)$ such that

$$|\Phi^{(n)}(x)| < k^n (2n)!, \qquad n \geqq 1,$$

or more generally

$$|\Phi^{(n)}(x)| < k^n \Gamma(\alpha n), \qquad n \geqq 1,$$

where $\alpha$ is a constant greater than unity. These functions are in general not analytic. In fact, it is possible to construct a function [5] satisfying this last inequality in a closed bounded interval, and such that the function and all its derivatives are zero at one point of the interval, but which is not identically zero in the interval.

In the work to follow we shall denote by $I$ either a closed bounded interval $[a, b]$, or an infinite interval of one of the three forms: $(-\infty, b]$, $(-\infty, \infty)$, $[a, \infty)$. If $\{M_n\}$ is a sequence of positive constants, we shall denote by $C_{\{M_n\}}$ the class of all functions $f(x)$ i.d. in an interval $I$ and such that to each function there corresponds a positive constant $k$ with the property that in $I$

$$|f^{(n)}(x)| < k^n M_n, \qquad n \geqq 1.$$

The class of analytic functions and the two classes of Gevray mentioned above, each defined in a closed bounded interval, are respectively classes $C_{\{n!\}}$, $C_{\{(2n)!\}}$ and $C_{\{\Gamma(\alpha n)\}}$.

The class $C_{\{n!\}}$ has the property that each function belonging to it such that the function and its derivatives of all orders are zero at one point of the interval is identically zero. Or what is equivalent, two functions each belong-

Presented to the Society, September 10, 1942; received by the editors August 15, 1941.

(1) Numbers in brackets refer to the references listed in the bibliography given at the end of the paper.

(2) See, for instance, S. Mandelbrojt, Rice Institute Pamphlet, vol. 29, 1942.

ing to this class, and having the same value and the same derivatives of all orders at one point of the interval are identically equal in the interval. A class $C_{\{M_n\}}$ having this property is called *quasi-analytic*. Thus the classes of Gevray are not quasi-analytic.

The physical character of Gevray's work indicates the importance of knowing necessary and sufficient conditions on the sequence $\{M_n\}$ in order that the class $C_{\{M_n\}}$ will be quasi-analytic. This problem was proposed by Hadamard(³) in 1912. Denjoy [3] gave a sufficient condition and Carleman [2] generalized Denjoy's condition and gave the complete answer to Hadamard's question. The following is Carleman's theorem in a form given by Ostrowski [8]:

Let $T(r) = \text{l.u.b.}_{n\geq 1}(r^n/M_n)$. A necessary and sufficient condition that the class $C_{\{M_n\}}$ will be quasi-analytic is that $\int_1^\infty \log T(r)dr/r^2 = \infty$.

If $I$ is the interval $0 \leq x \leq 2\pi$ and $C^*_{\{M_n\}}$ is the class of functions even and periodic with period $2\pi$ and belonging to $C_{\{M_n\}}$ in $I$, the class $C_{\{M_n\}}$ and the class $C^*_{\{M_n\}}$ are quasi-analytic at the same time. In other words, if there exists a function $f(x)$ not identically zero belonging to $C_{\{M_n\}}$ in $I$ such that the function and all its derivatives are zero at a point of $I$, then there exists a function $\Phi(x)$ not identically zero belonging to $C^*_{\{M_n\}}$ such that the function and all its derivatives are zero at a point of $I$.

It is known [4, 5] that if there exists in [0, 1] a function $f(x)$ not identically zero and belonging to $C_{\{M_n\}}$, with $f^{(n)}(0)=0$ $(n \geq 0)$, then there exists in the same interval a function $\Phi(x)$ not identically zero and belonging to $C_{\{M_n\}}$, with $\Phi^{(n)}(0)=\Phi^{(n)}(1)=0$ $(n\geq 0)$, and such that $\Phi(x)=\Phi(1-x)$. As a matter of fact, if $C_{\{M_n\}}$ is not quasi-analytic, there exists a function $\Phi(x)$ not identically zero in [0, 1] having the prescribed properties and such that $|\Phi^{(n)}(x)| < M_n$ $(n \geq 1)$, [5]. Thus on writing $\psi(x)=\Phi(x/2\pi)$ and continuing this function by periodicity with period $2\pi$, one sees immediately that if $C_{\{M_n\}}$ is not quasi-analytic, there exists a function $\psi(x)$ belonging to $C^*_{\{M_n\}}$ and not identically zero, such that $\psi^{(n)}(0)=\psi^{(n)}(2\pi)=0$ $(n\geq 0)$ and such that $|\psi^{(n)}(x)| < M_n$ $(n \geq 1)$, $x$ in $[0, 2\pi]$; that is, for which the constant $k$ is unity.

It can be shown that if $C^*_{\{M_n\}}$ is not quasi-analytic then there exists a function $F(s)=\sum_{m=1}^\infty a_m/m^s$, $s=\sigma+it$, where this series converges absolutely in all the plane ($\sigma_A$ = abscissa of absolute convergence $=-\infty$), the function $F(s)$ having the following properties:

(i) If $M(\sigma)=\sum_{m=1}^\infty |a_m|/m^\sigma$, then $M(-n) \leq M$ , $n=1, 2, \cdots$.

(ii) $F(-2n)=0$ for $n>n_0>0$.

(iii) $F(s)$ is not identically zero.

Indeed, if $C^*_{\{M_n\}}$ is not quasi-analytic, there exists an i.d. function $\Phi(x)=\sum_{m=0}^\infty d_m \cos mx$, not identically zero, with $\Phi^{(n)}(0)=\Phi^{(n)}(2\pi)=0$ $(n\geq 0)$ and such that $|\Phi^{(n)}(x)| < M_n$ $(n \geq 1)$. We then have

(³) J. Hadamard in a communication to Société Mathématique de France (Comptes Rendus des Séances de la Société Mathématique de France, 1912, p. 28).

$$\Phi^{(2p)}(0) = (-1)^p \sum_{m=1}^{\infty} d_m m^{2p} = 0, \qquad p \geqq 1. \tag{1}$$

But

$$d_m = \frac{1}{\pi} \int_0^{2\pi} \Phi(x) \cos mx \, dx,$$

and so after $q$ integrations by parts we have

$$d_m = \pm \frac{1}{\pi m^q} \int_0^{2\pi} \Phi^{(q)}(x) \begin{matrix} \cos mx \\ \sin mx \end{matrix} dx,$$

so that

$$| d_m | \leqq \frac{2M_q}{m^q}, \qquad q \geqq 1.$$

Therefore(4)

$$| d_m | \leqq \frac{2}{\text{l.u.b.}_{q\geqq 1} (m^q/M_q)} = \frac{2}{T(m)}.$$

Now put $a_m = d_m/2cm^2$ $(m \geqq 1)$ where $c = \sum_{m=1}^{\infty} 1/m^2$. Then the function $F(s) = \sum_{m=1}^{\infty} a_m/m^s$ has the desired properties. In the first place

$$M(-n) = \sum_{m=1}^{\infty} | a_m | m^n = \frac{1}{c} \sum_{m=1}^{\infty} \frac{| d_m | m^n}{2m^2} \leqq \frac{1}{c} \sum_{m=1}^{\infty} \frac{m^{n-2}}{T(m)}.$$

But $T(m) \geqq m^n/M_n$, and so

$$M(-n) \leqq \frac{M_n}{c} \sum_{m=1}^{\infty} 1/m^2 = M_n.$$

From (1) it follows that, for $n \geqq 2$:

$$F(-2n) = \sum_{m=1}^{\infty} a_m m^{2n} = \frac{1}{2c} \sum_{m=1}^{\infty} d_m m^{2(n-1)} = \frac{(-1)^{n-1}}{2c} \Phi^{[2(n-1)]}(0) = 0.$$

Finally, since $\Phi(x)$ is not identically zero, and since $\Phi(0) = 0$, the $d_m$ $(m \geqq 1)$ are not all zero and, all the $a_m$ are not zero, so that $F(s)$ is not identically zero.

The problem of finding conditions on the sequence $\{M_n\}$ in order that the function $F(s) = \sum_{m=1}^{\infty} a_m/m^s$ will be zero when $F(-2n) = 0$ $(n > n_0 > 0)$ and $\sum_{m=1}^{\infty} | a_m | m^n < M_n$, is analogous to, but technically very different from, problems considered by Carlson, Vl. Bernstein [1] and others, concerning entire functions $f(z)$, $z = x + iy$, or functions holomorphic in an angle, and having the following properties (considering only entire functions): If $m(r)$ is a function of $r$ increasing to infinity with $r$, $|f(z)| < m(r)$ for $|z| = r$ and $f(n) = 0$ $(n \geqq 1)$. What then are conditions on the function $m(r)$ in order that $f(z)$ will be identically zero? This problem was solved in many circumstances [1], and the solutions show that the requirement $f(n) = 0$ can be replaced by $f(\nu_n) = 0$, if the

(4) Such inequalities were first used by Mandelbrojt. See [4 or 5].

$|\nu_n|$ increase in a suitable manner. In the light of these results it seems natural to suppose that with the same conditions on the sequence $\{M_n\}$ in the problem concerning the series $\sum_{m=1}^{\infty} a_m/m^s$, $F(s)$ will be identically zero if we suppose $F(-\nu_n)=0$, the $\nu_n$ being real and tending toward infinity, with suitable hypotheses on the mode of increase of the $\nu_n$.

It seems reasonable then to expect that if the sequence $\{M_n\}$ satisfies the conditions of quasi-analyticity, or a condition analogous to these, a function belonging to $C^*_{\{M_n\}}$ will be zero when the function and its derivatives of orders $\nu_n$ are zero at a point, provided the integers $\nu_n$ increase in a suitable manner.

It is important to notice that while the conditions on the sequence $\{M_n\}$ are the same in order that a function in either the class $C_{\{M_n\}}$ or the class $C^*_{\{M_n\}}$ will be zero when it and its derivatives of *all* orders are zero at a point, it is impossible to extend to the class $C_{\{M_n\}}$ the notion of quasi-analyticity when only those derivatives of *certain orders* are zero at a point. The very simple example of the function $f(x)\equiv x$, which belongs to every class and which has derivatives of orders $n\geqq 2$ zero, proves it. This explains why we speak above only of the class $C^*_{\{M_n\}}$. Nevertheless, it will also be possible to extend our results to non-periodic functions of a class $C_{\{M_n\}}$: if the function behaves in a certain manner at infinity.

In a word, we are going to give conditions on the sequence $\{M_n\}$ in order that a function of $C^*_{\{M_n\}}$ will be identically zero knowing only that a suitable *partial sequence* of the derivatives of the function, along with the function itself, is zero at a point. In addition, we shall give conditions on the sequence $\{M_n\}$ in order that a function $f(x)$ of $C_{\{M_n\}}$ in $[0, \infty)$, such that $\int_0^\infty |f(x)|\,dx < \infty$, will be identically zero, again knowing only that a suitable *partial sequence* of its derivatives, along with the function itself, is zero at a point.

If we still suppose that a suitable partial sequence of its derivatives, along with the function itself, is zero at a point, but we no longer suppose that $\int_0^\infty |f(x)|\,dx<\infty$, we shall prove that if $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n}<\infty$, then from our conditions on $\{M_n\}$ it follows only that $\lim_{n\to\infty} |f^{(n)}(0)|^{1/n}=0$.

If the sequence of the orders of the derivatives supposed zero at a point is $\{\nu_m\}$, the distribution of the integers $\nu_m$ will be characterized by the quantity $G=\limsup_{m\to\infty} (\nu_m/m)$. In all cases our results will be valid if only $G<2$.

## PART I

1. It is well known that the classical problem of the quasi-analyticity of $C_{\{M_n\}}$ is equivalent to the problem of Watson (see [2, 4, 8]); that is, the problem of determining necessary and sufficient conditions on the sequence $\{M_n\}$ in order that a function $F(z)$, holomorphic in a half-plane (say $\Re(z)>1$[6]) and satisfying the inequality $|F(z)|<M_n/|z|^n$, will be identically zero. Our method consists in the very generalization of Watson's problem and the solution of the problem thus generalized.

(6) $\Re(z)$ denotes the real part of $z$.

We shall define analytic functions $F_1(z)$ and $F_2(z)$ as follows: If $f(x)$ belongs to $C^*_{\{M_n\}}$,

$$F_1(z) = \int_0^{2\pi} f(x)e^{-zx}dx.$$

If $f(x)$ belongs to $C_{\{M_n\}}$, where the interval of definition is taken to be $0 \leqq x < \infty$, and $f(x)$ is such that $\int_0^\infty |f(x)|dx < \infty$, we define

$$F_2(z) = \int_0^\infty f(x)e^{-zx}dx.$$

The function $F_1(z)$ is an entire function and $F_2(z)$ is holomorphic in the right half-plane $\mathfrak{R}(z) > 0$, and bounded there. That is, if $z$ is any point of $\mathfrak{R}(z) > 0$, $|F_2(z)|$ is less than the constant $\int_0^\infty |f(x)|dx$. If we integrate $n$ times by parts, we have

$$F_1(z) = (1 - e^{-2\pi z})\sum_{m=0}^{n-1} \frac{f^{(m)}(0)}{z^{m+1}} + \frac{1}{z^n}\int_0^{2\pi} f^{(n)}(x)e^{-zx}dx, \qquad \text{for all } z, \tag{2}$$

$$F_2(z) = \sum_{m=0}^{n-1} \frac{f^{(m)}(0)}{z^{m+1}} + \frac{1}{z^n}\int_0^\infty f^{(n)}(x)e^{-zx}dx, \qquad \mathfrak{R}(z) > 0. \tag{3}$$

In the classical case, where all derivatives of $f(x)$ are supposed zero at the origin, the sums appearing in (2) and (3) are zero, which enables one to pass directly to Watson's problem. In that case, for any $\alpha > 0$, in the right half-plane $\mathfrak{R}(z) \geqq \alpha$ both $|F_1(z)|$ and $|F_2(z)|$ are less than $ck^nM_n/|z|^n$, where $c$ is a constant and $k$ depends only on $f(x)$. This is not true, however, in the case when it is only known that a suitable *partial sequence* of the derivatives of $f(x)$ vanish at the origin.

Along with the hypothesis that the derivatives of $f(x)$ of orders $\nu_n$ are zero at the origin, we shall first assume that $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n} < \infty$. This restriction will be removed later.

While, from what was stated in the introduction, it seemed natural to conceive of our problem for the class $C^*_{\{M_n\}}$, composed of all even periodic functions belonging to $C_{\{M_n\}}$, and this problem certainly does not have meaning for all functions of this latter class, it will be expedient to begin our study for those functions of $C_{\{M_n\}}$ for which such a study is possible without supposing periodicity. The periodic case will be treated in Part II.

Let us first prove the following simple lemma:

LEMMA I. *Let $\{\nu_n\}$ be a sequence of positive integers, and $\{\lambda_n\}$ the sequence of integers complementary to the sequence $\{\nu_n\}$ with respect to the non-negative integers. If*

$$\liminf_{n\to\infty} \frac{\lambda_n}{n} \geqq G' > 1, \tag{a}$$

*then*

$$\text{(b)} \qquad \limsup_{n\to\infty} \frac{\nu_n}{n} \leqq \frac{G'}{G'-1}\cdot$$

*Conversely, from* (b), *with* $G'>1$, *follows* (a).

Denote by $M(t)$ the number of integers $\nu_n \leqq t$, and by $N(t)$ the number of integers $\lambda_n \leqq t$. If $\lambda_n \leqq t < \lambda_{n+1}$, then, for $t \geqq \lambda_1$

$$\frac{\lambda_n}{n} \leqq \frac{t}{N(t)} < \frac{\lambda_{n+1}}{n},$$

from which it follows that $\liminf_{t\to\infty} (t/N(t)) = \liminf_{n\to\infty} (\lambda_n/n)$. Thus from (a) we have that $\limsup_{t\to\infty}(N(t)/t) \leqq 1/G'$. On the other hand, $M(t)+N(t)=[t]+1$, where $[t]$ is the greatest integer contained in $t$. But from this it follows that

$$\liminf_{t\to\infty} \frac{M(t)}{t} + \limsup_{t\to\infty} \frac{N(t)}{t} = 1.$$

Since $\limsup_{t\to\infty} (N(t)/t) \leqq 1/G'$, we have that $\liminf_{t\to\infty} (M(t)/t) \geqq 1 - (1/G') = (G'-1)/G'$, and (b) follows from this.

The converse also follows immediately from the above considerations.

2. THEOREM I. *Suppose* $f(x)$ *belongs to* $C_{\{M_n\}}$ *in the interval* $0 \leqq x < \infty$ *and is such that* $\int_0^\infty |f(x)|\,dx < \infty$. *Moreover, suppose* $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n} < \infty$ *and* $f^{(\nu_m)}(0)=0$ $(m \geqq 1)$. *If* $\limsup_{m\to\infty} (\nu_m/m) < 2$, *and if* $\int_1^\infty \log T(r)dr/r^2 = \infty$, *where* $T(r) = \text{l.u.b.}_{n\geqq 1} (r^n/M_n)$, *then* $f(x)$ *is identically zero.*

If we denote by $R$ $(<\infty)$ the number $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n}$ and let $P_\beta = R+\beta$, where $\beta$ is any positive number, the series $\sum_{m=0}^\infty f^{(m)}(0)/z^{m+1}$ converges uniformly in $|z| \geqq P_\beta$ to a holomorphic function $\Phi(z)$. Let $\gamma$ be any positive number and denote by $D(\beta, \gamma)$ the region common to $|z| \geqq P_\beta$ and $\Re(z) \geqq \gamma$. Then in $D(\beta, \gamma)$, from (3) we have

$$\text{(4)} \qquad F_2(z) - \Phi(z) = \frac{1}{z^n}\int_0^\infty f^{(n)}(x)e^{-zx}dx - \sum_{m=n}^\infty \frac{f^{(m)}(0)}{z^{m+1}}\cdot$$

Since $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n} = R$, there is a constant $P > \max(R, 1)$ such that $|f^{(n)}(0)| < P^n$ for all $n \geqq 1$. Now let $\beta$ be chosen so that $P_\beta > P > 1$. Then in $|z| \geqq P_\beta$

$$\text{(5)} \qquad \left| \sum_{m=n}^\infty \frac{f^{(m)}(0)}{z^{m+1}} \right| \leqq \frac{P^n}{|z|^n} \sum_{m=0}^\infty \frac{P^m}{P_\beta^{m+1}} = \frac{1}{P_\beta - P}\,\frac{P^n}{|z|^n} = \frac{\kappa_1 P^n}{|z|^n},$$

where $\kappa_1$ is a constant.

On the other hand, $f(x)$ belongs to $C_{\{M_n\}}$ in $0 \leqq x < \infty$ and so $|f^{(n)}(x)|$

$<k^n M_n$ $(n \geqq 1)$ for $0 \leqq x < \infty$, where $k$ depends only on $f(x)$, from which it follows that

$$\left| \frac{1}{z^n} \int_0^\infty f^{(n)}(x) e^{-zx} dx \right| < \frac{k^n M_n}{|z|^n} \int_0^\infty |e^{-zx}| dx.$$

Hence for all $z$ in $\Re(z) \geqq \gamma$,

$$\left| \frac{1}{z^n} \int_0^\infty f^{(n)}(x) e^{-zx} dx \right| < \frac{k^n M_n}{|z|^n} \int_0^\infty e^{-\gamma x} dx = \frac{\kappa_2 k^n M_n}{|z|^n}, \qquad n \geqq 1, \tag{6}$$

where $\kappa_2$ is a constant.

From (4), (5) and (6), it results that in $D(\beta, \gamma)$

$$|F_2(z) - \Phi(z)| < \frac{\kappa_1 P^n + \kappa_2 k^n M_n}{|z|^n}, \qquad n \geqq 1.$$

If $\liminf_{n \to \infty} (M_n)^{1/n} > 0$, there is a constant $c$ such that

$$|F_2(z) - \Phi(z)| < \frac{c^n M_n}{|z|^n}, \qquad n \geqq 1,$$

for all $z$ in $D(\beta, \gamma)$. But since $\int_1^\infty \log T(r)\, dr/r^2 = \infty$, by Carleman's solution of Watson's problem [4, 8], $F_2(z) \equiv \Phi(z)$ in $D(\beta, \gamma)$.

If $\liminf_{n \to \infty} (M_n)^{1/n} = 0$, to any positive constant $q$ $(< \infty)$ there corresponds a sequence $\{n_i\}$ of integers such that $M_{n_i}^{1/n_i} < q$ $(i \geqq 1)$. Therefore in $D(\beta, \gamma)$

$$|F_2(z) - \Phi(z)| < \frac{\kappa_1 P^{n_i} + \kappa_2 k^{n_i} q^{n_i}}{|z|^{n_i}} < \frac{Q^{n_i}}{|z|^{n_i}}, \qquad i \geqq 1,$$

where $Q$ $(< \infty)$ is a constant. If we denote by $\{M_n'\}$ the sequence such that $M_{n_i}' = Q^{n_i}$ $(i \geqq 1)$ and $M_n' = \kappa_1 P^n + \kappa_2 k^n M_n$, $n$ not equal to any $n_i$, then in $D(\beta, \gamma)$

$$|F_2(z) - \Phi(z)| < \frac{M_n'}{|z|^n}.$$

Now if we let $T_1(r) = \text{l.u.b.}_{n \geqq 1} (r^n / M_n')$, then $T_1(r) = \infty$ for $r > Q$. For, if $r > Q$, we have

$$T_1(r) \geqq \text{l.u.b.}_{i \geqq 1} \frac{r^{n_i}}{M_{n_i}'} = \text{l.u.b.}_{i \geqq 1} (r/Q)^{n_i} = \infty.$$

Therefore $\int_1^\infty \log T_1(r)\, dr/r^2 = \infty$ and again $F_2(z) \equiv \Phi(z)$ in $D(\beta, \gamma)$.

Thus, under the hypotheses of the theorem, we have in all cases that $F_2(z) \equiv \Phi(z)$ in $D(\beta, \gamma)$ and hence in the entire region of definition of the functions.

If we let $\psi(\zeta)=F_2(1/\zeta)$, $\psi(\zeta)$ is holomorphic in $\Re(\zeta)>0$, and in $|\zeta|<1/R$,

$$\psi(\zeta) = \sum_{m=0}^{\infty} f^{(m)}(0)\zeta^{m+1} = \zeta\sum_{m=1}^{\infty} f^{(\lambda_m)}(0)\zeta^{\lambda_m} = \zeta\sum_{m=1}^{\infty} a_m\zeta^{\lambda_m}, \tag{7}$$

where $\{\lambda_m\}$ is the sequence of integers complementary to the sequence $\{\nu_m\}$ with respect to the non-negative integers. By hypotheses $\int_0^\infty |f(x)|\,dx<\infty$ and so $|F_2(s)|\leqq\int_0^\infty|f(x)|\,dx$ for all $s$ in $\Re(s)>0$. Therefore we have that *$\psi(\zeta)$ is holomorphic and bounded in $\Re(\zeta)>0$ and is given by (7) in $|\zeta|<1/R$.*

The conclusion of the theorem is now obtained from the following fact, which for later reference will be stated as

Lemma II. *Let the function $F(z)$ be defined by the series $\sum_{m=1}^{\infty}a_mz^{\lambda_m}$ in $|z|<\rho$ and let $G'=\liminf_{m\to\infty}(\lambda_m/m)$. If $F(z)$ is not identically a constant, in each closed angle with vertex at the origin and with opening $2\pi/G'$, one of the two following possibilities must exist. Either*

(i) *$F(z)$ has a singular point, or*

(ii) *$F(z)$ is not bounded.*

This result is a part of a theorem of Mandelbrojt. For the proof see [6' p. 15 ff.].

Since $\{\nu_m\}$ and $\{\lambda_m\}$ are complementary sequences with respect to the non-negative integers, and since $\limsup_{m\to\infty}(\nu_m/m)<2$, it follows from Lemma I that $G'>2$. Hence there exist closed angles in the right half-plane with vertex at the origin and with opening $2\pi/G'$. Therefore $\psi(\zeta)$ must be identically a constant since neither (i) nor (ii) is true for $\psi(\zeta)$ in the right half-plane. Moreover, since $\psi(0)=0$, this constant must be zero. Therefore $a_m=f^{(\lambda_m)}(0)=0$ for all $m\geqq 1$. Thus all derivatives of $f(x)$, along with $f(x)$ itself, are zero at $x=0$ since $f^{(\nu_m)}(0)=0$, $m\geqq 1$. Then, since $\int_1^\infty\log T(r)\,dr/r^2=\infty$, the class $C_{\{M_n\}}$ is quasi-analytic and $f(x)$ is identically zero.

3. If the restriction $\int_0^\infty|f(x)|\,dx<\infty$ is removed, the following result can be obtained:

Theorem II. *Suppose $f(x)$ belongs to $C_{\{M_n\}}$ in $0\leqq x<\infty$ and $f^{(\nu_m)}(0)=0$' $m\geqq 1$. Moreover, suppose* $\limsup_{n\to\infty}|f^{(n)}(0)|^{1/n}<\infty$. *If* $\limsup_{m\to\infty}(\nu_m/m)<2$' *and if $\int_1^\infty\log T(r)\,dr/r^2=\infty$, then* $\lim_{n\to\infty}|f^{(n)}(0)|^{1/n}=0$.

In order to prove this theorem we shall use the following lemmas.

Lemma III. *Let $\theta(s)$ be defined by the Dirichlet series $\sum_{m=1}^{\infty}d_me^{-\mu_ms}$, where* $\liminf_{m\to\infty}\mu_m/m\geqq L>0$. *Let $\epsilon$ be a positive quantity and $s_1=\sigma_1+it_1$. If $\theta(s)$ is holomorphic in the circle $|s-s_1|\leqq(\pi/L)(1+\epsilon)$*(6) *and if $|\theta(s)|<M$ in this circle*

(6) A function represented by a Dirichlet series, which, of course, we suppose always to have an axis of convergence, will be said to be holomorphic at a point $\sigma_0+it_0$, if it is possible to continue analytically the function given by the sum of the series along the line $t=t_0$ through the point $\sigma_0+it_0$. The value $\theta(\sigma_0+it_0)$ is the value given by this continuation.

*then for each* $j=1, 2, \cdots,$

$$|d_j|\,|g_j(\mu_j)|\,e^{-\mu_j\sigma_1} < K_\epsilon M,$$

*where* $K_\epsilon$ *is independent of* $j$, $\sigma_1$ *and* $M$, *and*

$$g_j(z) = z \prod_{n=1, n\neq j}^{\infty}{}' (1 - z^2/\mu_n^2)\text{(7)}. \tag{8}$$

If $\theta(s)$ is bounded in the whole strip $|t| < \pi/2$, and if $L > 2$, since $\mu_1 > 0$ it is then evident that $d_j = 0$ for every $j = 1, 2, \cdots$.

This lemma was proved by Mandelbrojt [6, p. 14 ff.] and was used by him in the proof of the theorem which is given in the present paper in a modified form as Lemma II. Lemma III was also proved for entire functions by Mandelbrojt and Gergen [7].

From Lemma III results the following lemma which was given by Mandelbrojt in his lectures at Rice Institute and for which we shall now give the proof:

LEMMA IV. *Let* $F(z)$ *be given in the neighborhood of the origin by the Taylor series* $\sum_{m=1}^{\infty} a_m z^{\mu_m}$, *with* $\liminf_{m\to\infty} (\mu_m/m) = D > L$. *If* $F(z)$ *is holomorphic in the sector* $|z| < \rho e^{\pi/L}$, $|\arg z| < \pi/L$, *then* $\limsup_{m\to\infty} |a_m|^{1/\mu_m} \leqq 2^3/\rho$(8).

If we put $z = e^{-s}$, the function $\Phi(s) = F(e^{-s})$ is holomorphic for values of $s = \sigma + it$ such that $\sigma > -\log \rho - \pi/L$, $|t| < \pi/L$, and in a half-plane $\sigma > c$, $\Phi(s)$ is given by the series $\sum_{m=1}^{\infty} a_m e^{-\mu_m s}$.

There exists therefore a constant $\epsilon > 0$ such that $\Phi(s)$ is holomorphic in the circle $|s - s_0| \leqq (\pi/D)(1+\epsilon)$, where $s_0 = -\log \rho$. Hence by Lemma III we have

$$|a_j|\,|g_j(\mu_j)| < K_\epsilon M \rho^{-\mu_j}, \tag{9}$$

where $g_j(z)$ is again the function defined in (8) and $M = \max |\Phi(z)|$ in

$$|s - s_0| \leqq (\pi/D)(1+\epsilon).$$

We now evaluate $g_j(\mu_j)$. Since the $\mu_m$ are increasing positive integers.

$$\mu_{j+k} \geqq \mu_j + k, \qquad k > 0,$$

and

$$\mu_{j-k} \leqq \mu_j - k, \qquad 0 \leqq k < j.$$

Therefore

(7) We suppose $\mu_1 > 0$, and as usual, to $\mu_n$ increasing to infinity.

(8) We again suppose $\mu_1 > 0$.

$$
\begin{aligned}
|g_j(\mu_j)| &= \mu_j \left| \prod_{n=1, n\neq j}^{\infty}{}' (1 - \mu_j^2/\mu_n^2) \right| \geqq \mu_j \prod_{p=-j+1, p\neq 0}^{\infty}{}' \left| 1 - \frac{\mu_j^2}{(\mu_j + p)^2} \right| \\
&= \mu_j \prod_{m=\mu_j-j+1, m\neq\mu_j}^{\infty}{}' \left| 1 - \mu_j^2/m^2 \right| \\
&= \mu_j \left[ \prod_{m=1, m\neq\mu_j}^{\infty}{}' \left| 1 - \mu_j^2/m^2 \right| \right] \left[ \prod_{m=1}^{\mu_j - j} \left| 1 - \mu_j^2/m^2 \right| \right]^{-1} \text{(9)}.
\end{aligned} \tag{10}
$$

But since

$$
\begin{aligned}
\mu_j \prod_{m=1, m\neq\mu_j}^{\infty}{}' (1 - \mu_j^2/m^2) &= \lim_{z\to\mu_j} z \prod_{m=1}^{\infty} (1 - z^2/m^2) \cdot \frac{\mu_j^2}{\mu_j^2 - z^2} \\
&= \lim_{z\to\mu_j} \frac{\sin \pi z}{\pi} \cdot \frac{\mu_j^2}{\mu_j^2 - z^2},
\end{aligned}
$$

we see that the left member of this last equality is equal to $-\mu_j \cos \pi\mu_j/2 = (-1)^{\mu_j - 1}\mu_j/2$.

On the other hand,

$$
\begin{aligned}
\prod_{m=1}^{\mu_j - j} \left| 1 - \mu_j^2/m^2 \right| &= \left[ \prod_{m=1}^{\mu_j - j} \frac{\mu_j - m}{m} \right] \left[ \prod_{m=1}^{\mu_j - j} \frac{\mu_j + m}{m} \right] \\
&= \frac{(\mu_j - 1)!}{(j-1)!(\mu_j - j)!} \frac{(2\mu_j - j)!}{\mu_j!(\mu_j - j)!} = \frac{j}{\mu_j} \frac{(2\mu_j - j)!}{[(\mu_j - j)!]^2 j!} \\
&= \frac{j}{\mu_j} \frac{(2\mu_j - j)!}{\mu_j!(\mu_j - j)!} \cdot \frac{\mu_j!}{(\mu_j - j)! j!} = \frac{j}{\mu_j} C_{2\mu_j - j}^{\mu_j} C_{\mu_j}^{j} \\
&\leqq \frac{j}{\mu_j} 2^{2\mu_j - j} 2^{\mu_j} = \frac{j}{\mu_j} 2^{3\mu_j - j}.
\end{aligned}
$$

From (10) we then have that

$$
|g_j(\mu_j)| \geqq (\mu_j^2/j) 2^{j - 3\mu_j - 1}.
$$

From (9) it then follows that

$$
|a_j| \leqq K_\epsilon M (j/\mu_j^2) 2^{3\mu_j - j - 1} \rho^{-\mu_j},
$$

and from this the statement of the lemma results immediately.

We are now ready to pass to the proof of Theorem II. It should be noticed that even under the present hypotheses, the function $F_2(z)$ defined by (3) exists for each $z$ in $\mathfrak{R}(z) > 0$ and is holomorphic there. For,

---

(9) If $\mu_j = j$, the quantity in the second brackets must be taken equal to unity wherever it appears.

$$f(x) = f(0) + \int_0^x f'(t)dt,$$

and $|f'(x)| < kM_1$ in $0 \leqq x < \infty$, so that $|f(x)| \leqq |f(0)| + kM_1x$, $0 \leqq x < \infty$. Therefore $\int_0^\infty f(x)e^{-zx}dx$ converges uniformly in $\Re(z) \geqq \alpha > 0$ and defines a holomorphic function there. But under the present hypotheses it can no longer be said that $F_2(z)$ is bounded in $\Re(z) > 0$ and so it is not possible to draw the same conclusions as in Theorem I. However, the function $\psi(\zeta) = F_2(1/\zeta)$ is holomorphic in $\Re(\zeta) > 0$ and is given by the series $\zeta\sum_{m=1}^\infty f^{(\lambda_m)}(0)\zeta^{\lambda_m}$ in $|\zeta| < 1/R$ where $R = \limsup_{n\to\infty} |f^{(n)}(0)|^{1/n}$. By Lemma I, $\liminf_{m\to\infty}(\lambda_m/m) > 2$ since $\limsup_{m\to\infty} (\nu_m/m) < 2$. Hence we can apply Lemma IV with $L = 2$ to the function $\psi(\zeta)$. The sequence $\{\mu_m\}$ is replaced by the sequence $\{\lambda_m + 1\}$, $a_m = f^{(\lambda_m)}(0)$ and $\rho$ can be any positive number. Therefore

$$\limsup_{m\to\infty} |f^{(\lambda_m)}(0)|^{1/\lambda_m} \leqq 2^3/\rho,$$

and since $\rho$ is any positive number

$$\lim_{n\to\infty} |f^{(n)}(0)|^{1/n} = \lim_{m\to\infty} |f^{(\lambda_m)}(0)|^{1/\lambda_m} = 0.$$

4. We now pass to the case in which it is not assumed that $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n} < \infty$. In this case the radius of convergence of the series $\sum_{m=1}^\infty f^{(\lambda_m)}(0)\zeta^{\lambda_m+1}$ may be zero and then the series no longer represents $\psi(\zeta)$. But we shall see that due to the relation which exists between this series and the function, it is possible to consider the series as somewhat analogous to an asymptotic development in order to obtain some properties of $\psi(\zeta)$.

It is useful first to establish some preliminary lemmas.

LEMMA V. *If $\Phi(z)$ is not identically zero and is holomorphic in $|z| \leqq 1$ except possibly at $z = -1$, and is bounded in this circle, then*

$$\int_{-\pi}^{\pi} \log |\Phi(e^{i\theta})| d\theta > -\infty.$$

This lemma is well known. A proof is given by Ostrowski [8].

Let $\{N_n\}$ be a sequence of positive numbers and $\{\mu_n\}$ a sequence of positive numbers increasing to infinity. If $\omega$ is any positive number, we define

$$\tau_\omega(r) = \underset{n\geqq 1}{\text{l.u.b.}} \frac{r^{\omega\mu_n}}{N_n}.$$

LEMMA VI. *Let $\{N_n\}$ be a sequence of positive numbers and $\{\mu_n\}$ a sequence of positive numbers increasing to infinity. Suppose $A(\xi)$ is holomorphic and bounded in $\Re(\xi) \geqq 0$, and*

(11) $$| A(\xi) | < \frac{k^{\omega\mu_n} N_n}{| \xi |^{\omega\mu_n}}, \qquad n \geqq 1, \text{ in } \mathfrak{R}(\xi) \geqq 0,$$

*where k is a constant. If* $\int_1^\infty \log \tau_\omega(r) dr/r^2 = \infty$, *then* $A(\xi) \equiv 0$.

From (11) we have

$$| A(\xi) | \leqq \frac{1}{\text{l.u.b.}_{n\geq 1} [(| \xi | /k)^{\omega\mu_n}/N_n]} = \frac{1}{\tau_\omega(| \xi | /k)}, \qquad \mathfrak{R}(\xi) \geqq 0,$$

and hence

(12) $$\log | A(\xi) | \leqq - \log \tau_\omega(| \xi | /k).$$

Now let $\xi = (1-z)/(1+z)$ and $B(z) = A[(1-z)/(1+z)]$. $B(z)$ is then holomorphic in $|z| \leqq 1$ except possibly at $z = -1$, and is bounded in this circle. If $A(\xi)$ were not identically zero, $B(z)$ would not be identically zero and by Lemma V we would have

$$\int_{-\pi}^{\pi} \log | B(e^{i\theta}) | d\theta = \int_{-\pi}^{\pi} \log \left| A\left(\frac{1 - e^{i\theta}}{1 + e^{i\theta}}\right) \right| d\theta > -\infty.$$

It would then follow from (12) that

$$2 \int_0^{\pi} \log \tau_\omega\left(\frac{| 1 - e^{i\theta} |}{k | 1 + e^{i\theta} |}\right) d\theta = \int_{-\pi}^{\pi} \log \tau_\omega\left(\frac{| 1 - e^{i\theta} |}{k | 1 + e^{i\theta} |}\right) d\theta$$

$$\leqq - \int_{-\pi}^{\pi} \log \left| A\left(\frac{1 - e^{i\theta}}{1 + e^{i\theta}}\right) \right| d\theta < \infty.$$

But on letting $(1/k) | (1-e^{i\theta})/(1+e^{i\theta}) | = (1/k) \tan (\theta/2) = r$ we see that the first integral is equal to

$$2k \int_0^\infty \frac{\log \tau_\omega(r)}{1 + r^2k^2} dr,$$

and since $\lim_{r\to\infty} [r^2/(1+r^2k^2)] = 1/k^2$, we would have

$$\int_1^\infty \frac{\log \tau_\omega(r)}{r^2} dr < \infty,$$

which is contrary to hypotheses. Hence it must be true that $A(\xi) \equiv 0$.

We shall now prove the following lemma concerning series which are analogous to an asymptotic series. This lemma is a generalization of the theorem of Mandelbrojt which is given in the present paper as Lemma III.

Lemma VII. *If the sequence* $\{d_n\}$ *is such that there exist a sequence* $\{N_n\}$ *of positive numbers, a sequence* $\{\mu_n\}$ *of increasing positive numbers with* $\liminf_{n\to\infty} (\mu_n/n) = G' > 2$, *a positive quantity* $\omega < 1 - 2/G'$, *the two sequences and*

*the quantity* $\omega$ *related by* $\int_1^\infty \log \tau_\omega(r)dr/r^2 = \infty$, *and a function* $\theta(s)$, $s=\sigma+it$, *holomorphic and bounded for* $|t|<\pi/2$, *satisfying for every* $0<\delta<1$ *the inequalities*

$$\left|\theta(s) - \sum_{m=1}^{n} d_m e^{-\mu_m s}\right| < L_\delta N_n e^{-\mu_n\sigma}, \qquad n \geqq 1,\ |t| < \pi(1-\delta)/2,\ \sigma > \sigma_\delta,$$

*where* $L_\delta$ *and* $\sigma_\delta$ *depend only on* $\delta$, *then* $d_n = 0$ $(n \geqq 1)$.

The proof has some points in common with the proof of the theorem of Mandelbrojt given as Lemma III. The function $g_j(z)$ given by (8) is entire and odd. Hence its expansion in powers of $z$ contains only odd powers:

$$g_j(z) = \sum_{p=0}^{\infty} C_{2p+1}^{(j)} z^{2p+1}.$$

In the paper by Mandelbrojt and Gergen [7], it was proved that

$$|C_{2p+1}^{(j)}| < \frac{A_\epsilon}{(2p+1)!}\left[\frac{\pi}{G'}(1+\epsilon)\right]^{2p+1}, \tag{13}$$

where $A_\epsilon$ depends only on $\epsilon$.

Define the function $\theta_j(s)$ as follows:

$$\theta_j(s) = \sum_{p=0}^{\infty} C_{2p+1}^{(j)} \theta^{(2p+1)}(s).$$

Let $\epsilon>0$ and a positive $\delta<1$ be chosen so that $\omega' = 1-(2/G')(1+\epsilon)-\delta > \omega > 0$. We shall now prove that the series defining $\theta_j(s)$ converges uniformly and absolutely in the strip $S_\epsilon$: $|t| < \pi(\omega'+\delta)/2$. If $s$ is any point of $S_\epsilon$, each point $u$ of the circumference $\gamma$: $|u-s| = \pi(1+\epsilon)/G'$ is in the strip $|t|<\pi/2$. Then by Cauchy's integral formula,

$$|\theta^{(2p+1)}(s)| \leqq \frac{(2p+1)!}{2\pi}\int_\gamma \frac{|\theta(u)|\,|du|}{|u-s|^{2p+2}} \leqq \frac{(2p+1)!M}{[\pi(1+\epsilon)/G']^{2p+1}},$$

where $M$ is the bound on $\theta(s)$ in $|t|<\pi/2$. Therefore from this inequality and (13) with $\epsilon$ replaced by $\epsilon/2$, we have that in $S_\epsilon$

$$|C_{2p+1}^{(j)}\theta^{(2p+1)}(s)| \leqq MA_{\epsilon/2}\left[\frac{1+\epsilon/2}{1+\epsilon}\right]^{2p+1},$$

and so the series defining $\theta_j(s)$ converges uniformly and absolutely. Moreover it also follows from this last inequality that

$$|\theta_j(s)| \leqq MA_{\epsilon/2}\sum_{p=0}^{\infty}\left(\frac{1+\epsilon/2}{1+\epsilon}\right)^{2p+1} = K_\epsilon M, \tag{14}$$

for each $s$ in $S_\epsilon$, where $K_\epsilon$ depends only on $\epsilon$.

If we let

$$H_n(s) = \theta(s) - \sum_{m=1}^{n} d_m e^{-\mu_m s},$$

we have by hypothesis that in $|t| < \pi(1-\delta)/2$, $\sigma > \sigma_\delta$,

$$|H_n(s)| < L_\delta N_n e^{-\mu_n \sigma}, \qquad n \geqq 1.$$

Let $S_{\epsilon,\delta}$ be the half-strip defined by $|t| < \pi\omega'/2$, $\sigma > \sigma_\delta + \pi(1+\epsilon)/G' = \sigma'_\delta$. If $s$ is any point of the half-strip $S_{\epsilon,\delta}$, each point of the circumference $\gamma$ will be a point of the half-strip $|t| < \pi(1-\delta)/2$, $\sigma > \sigma_\delta$. Then by the Cauchy integral formula for $H_n^{(2p+1)}(s)$ we have that in $S_{\epsilon,\delta}$

$$\left| \theta^{(2p+1)}(s) + \sum_{m=1}^{n} d_m \mu_m^{2p+1} e^{-\mu_m s} \right| < \frac{(2p+1)!L_\delta N_n e^{-\mu_n \sigma}}{[\pi(1+\epsilon)/G']^{2p+1}}.$$

This inequality along with (13) yields that in $S_{\epsilon,\delta}$

$$(15)\qquad \left| C_{2p+1}^{(j)} \left[ \theta^{(2p+1)}(s) + \sum_{m=1}^{n} d_m \mu_m^{2p+1} e^{-\mu_m s} \right] \right| < A_{\epsilon/2} L_\delta N_n e^{-\mu_n \sigma} \left( \frac{1+\epsilon/2}{1+\epsilon} \right)^{2p+1}.$$

Now consider the series

$$\sum_{p=0}^{\infty} C_{2p+1}^{(j)} \left[ \theta^{(2p+1)}(s) + \sum_{m=1}^{n} d_m \mu_m^{2p+1} e^{-\mu_m s} \right].$$

By (15) this series converges uniformly and absolutely in $S_{\epsilon,\delta}$ and in this half-strip is equal to

$$\theta_j(s) + \sum_{m=1}^{n} d_m e^{-\mu_m s} \left( \sum_{p=0}^{\infty} C_{2p+1}^{(j)} \mu_m^{2p+1} \right) = \theta_j(s) + \sum_{m=1}^{n} d_m g_j(\mu_m) e^{-\mu_m s}.$$

But since $g_j(\mu_m) = 0$ for $m \neq j$, if $n \geqq j$, this last expression becomes $\theta_j(s) + d_j g_j(\mu_j) e^{-\mu_j s}$, and again from (15) it follows that in $S_{\epsilon,\delta}$

$$|\theta_j(s) + d_j g_j(\mu_j) e^{-\mu_j s}| \leqq A_{\epsilon/2} L_\delta N_n e^{-\mu_n \sigma} \sum_{p=0}^{\infty} \left( \frac{1+\epsilon/2}{1+\epsilon} \right)^{2p+1} = L_\delta K'_\epsilon N_n e^{-\mu_n \sigma}, \qquad n \geqq j,$$

where $K'_\epsilon$ depends only on $\epsilon$.

Let $\eta = \eta_1 + \eta_2 i = s/\omega'$. Then if we let $E_j(\eta) = \theta_j(\eta\omega')$, we have that in $\eta_1 > \sigma'_\delta/\omega'$, $|\eta_2| < \pi/2$

$$|E_j(\eta) + d_j g_j(\mu_j) e^{-\mu_j \omega' \eta}| < L_\delta K'_\epsilon N_n e^{-\mu_n \omega' \eta_1}, \qquad n \geqq j.$$

Now let $\xi = e^\eta$ and denote by $A_j(\xi)$ the function corresponding by this transformation to $E_j(\eta) + d_j g_j(\mu_j) e^{-\mu_j \omega' \eta}$. We then have in the domain defined by $\Re(\xi) > 0$, $|\xi| > e^{\sigma'_\delta/\omega'}$, since $|\xi| = e^{\eta_1}$,

$$|A_j(\xi)| < \frac{L_\delta K_\epsilon' N_n}{|\xi|^{\omega'\mu_n}}, \qquad n \geqq j.$$

But then there exists a constant $k>1$ such that for every $1 \leqq n < j$

$$|A_j(\xi)| < \frac{L_\delta K_\epsilon' k^{\omega'\mu_n} N_n}{|\xi|^{\omega'\mu_n}}$$

in $\Re(\xi)>0$, $|\xi|>e^{\sigma'\delta/\omega'}$. Thus we have that

$$|A_j(\xi)| < \frac{L_\delta K_\epsilon' k^{\omega'\mu_n} N_n}{|\xi|^{\omega'\mu_n}}, \qquad n \geqq 1,$$

in $\Re(\xi)>0$, $|\xi|>e^{\sigma'\delta/\omega'}$.

Now let

$$A(\xi) = A_j(a+\xi)/L_\delta K_\epsilon',$$

where $a=e^{\sigma'\delta/\omega'}+1$. $A(\xi)$ is holomorphic in $\Re(\xi)\geqq 0$ and from (14) it follows that it is bounded in this closed half-plane. From the above inequality we have that

$$|A(\xi)| = \frac{|A_j(a+\xi)|}{L_\delta K_\epsilon'} \leqq \frac{k^{\omega'\mu_n} N_n}{|\xi + a|^{\omega'\mu_n}} < \frac{k^{\omega'\mu_n} N_n}{|\xi|^{\omega'\mu_n}}, \qquad n \geqq 1,$$

in $\Re(\xi)\geqq 0$, since $a$ is a real positive quantity.

Since $\omega'>\omega$, if $\int_1^\infty \log \tau_\omega(r)dr/r^2=\infty$, then $\int_1^\infty \log \tau_{\omega'}(r)dr/r^2=\infty$. Thus all the conditions of Lemma VI are satisfied by the function $A(\xi)$ and the sequences $\{\mu_n\}$ and $\{N_n\}$, where the $\omega$ of that lemma becomes the present $\omega'$. Therefore $A(\xi)\equiv 0$, and we have that $A_j(\xi)\equiv 0$. From this it follows that

$$\theta_j(s) \equiv - d_j g_j(\mu_j) e^{-\mu_j s}.$$

From (14) we then have that

$$|d_j|\,|g_j(\mu_j)|\, e^{-\mu_j\sigma} < MK_\epsilon$$

for all $\sigma$. And since this inequality holds for all $\sigma$, and since $\mu_1>0$, it follows that $d_j=0$ for $j=1, 2, \cdots$. With this the lemma is proved.

Let there be given a sequence $\{M_n\}$ of positive numbers, a sequence $\{\nu_n\}$ of positive integers and a positive constant $\omega$. We shall define

$$T_\omega(r) = \operatorname*{l.u.b.}_{n\geqq 1} \frac{r^{\omega\lambda_n}}{M_{\lambda_n+1}},$$

where $\{\lambda_n\}$ is the sequence of positive integers complementary to the sequence $\{\nu_n\}$ with respect to the non-negative integers.

THEOREM III. *Suppose $f(x)$ belongs to $C_{\{M_n\}}$ in $0\leqq x<\infty$ and is such that $\int_0^\infty |f(x)|dx<\infty$. Moreover, suppose $f^{(\nu_n)}(0)=0$ $(n\geqq 1)$. If $G=\limsup_{n\to\infty}(\nu_n/n)$*

$<2$, *and if there exists a positive constant* $\omega<2/G-1$ *such that* $\int_0^\infty \log T_\omega(r)dr/r^2 = \infty$, *then* $f(x)$ *is identically zero.*

We shall again consider $F_2(z)$ defined by (3), but now in the half-plane $\Re(z)\geqq\alpha>0$. We then have in this half-plane

$$\left| F_2(z) - \sum_{m=0}^{n-1} \frac{f^{(m)}(0)}{z^{m+1}} \right| \leqq \frac{1}{|z|^n}\int_0^\infty |f^{(n)}(x)e^{-zx}|\, dx.$$

Since $f(x)$ belongs to $C_{\{M_n\}}$ in $0\leqq x<\infty$, it follows that

$$\left| F_2(z) - \sum_{m=0}^{n-1} \frac{f^{(m)}(0)}{z^{m+1}} \right| < \frac{ck^nM_n}{|z|^n}, \qquad n \geqq 1,$$

in $\Re(z)\geqq\alpha$, where $c$ is a constant.

If we let $z=e^s$, $s=\sigma+it$, and $\Psi(s)=F_2(e^s)$, on recalling that $f^{(\nu_n)}(0)=0$ $(n\geqq 1)$, the last inequality becomes

$$\left| \Psi(s) - \sum_{m=1}^{n} f^{(\lambda_m)}(0)e^{-(\lambda_m+1)s} \right| < ck^{\lambda_n+1}M_{\lambda_n+1}e^{-(\lambda_n+1)\sigma} \tag{16}$$

in $D_\alpha$, where $D_\alpha$ is the image of $\Re(z)\geqq\alpha$ by the transformation $z=e^s$. Here $\{\lambda_n\}$ is the sequence of integers complementary to the sequence $\{\nu_n\}$ with respect to the non-negative integers. Let $G'=\lim\inf_{n\to\infty}(\lambda_n/n)$. If we let $\mu_n=\lambda_n+1$, it is evident that $G'=\lim\inf_{n\to\infty}(\mu_n/n)$. Since by hypotheses $G<2$, we have by Lemma I that $G'>2$.

To each $0<\delta<1$, there corresponds a number $\sigma_\delta$ such that each point of the half-strip $|t|<\pi(1-\delta)/2$, $\sigma>\sigma_\delta$ is a point of $D_\alpha$. Therefore from (16) we have that in this half-strip

$$\left| \Psi(s) - \sum_{m=1}^{n} f^{(\lambda_m)}(0)e^{-\mu_m s} \right| < N_n e^{-\mu_n\sigma}, \qquad n \geqq 1,$$

where $N_n=ck^{\mu_n}M_{\mu_n}$.

We also have

$$\tau_\omega(r) = \underset{n\geqq 1}{\text{l.u.b.}}\frac{r^{\omega\mu_n}}{N_n} = \underset{n\geqq 1}{\text{l.u.b.}}\frac{r^{\omega(\lambda_n+1)}}{ck^{\lambda_n+1}M_{\lambda_n+1}} = \frac{r^\omega}{ck}\underset{n\geqq 1}{\text{l.u.b.}}\frac{r^{\omega\lambda_n}}{M_{\lambda_n+1}k^{\lambda_n}}$$
$$= \frac{r^\omega}{ck}T_\omega\left(\frac{r}{k^{1/\omega}}\right).$$

From this it is then evident that the divergence of $\int_1^\infty \log \tau_\omega(r)dr/r^2$ follows from the divergence of $\int_1^\infty \log T_\omega(r)dr/r^2$.

Thus $\Psi(s)$ satisfies all the conditions on the function $\theta(s)$ in Lemma VII, where $f^{(\lambda_n)}(0)=d_n$. Hence by this lemma $f^{(\lambda_n)}(0)=0$, $n\geqq 1$. But then all the derivatives of $f(x)$ as well as the function itself are zero at $x=0$.

On the other hand, since $\omega<2/G-1=1-2/G'<1$, if $r>1$:

$$T(r) = \underset{n\geqq 1}{\text{l.u.b.}} \frac{r^n}{M_n} \geqq \underset{n\geqq 1}{\text{l.u.b.}} \frac{r^{\lambda_n+1}}{M_{\lambda_n+1}} \geqq \underset{n\geqq 1}{\text{l.u.b.}} \frac{r^{\omega(\lambda_n+1)}}{M_{\lambda_n+1}} = r^\omega \underset{n\geqq 1}{\text{l.u.b.}} \frac{r^{\omega\lambda_n}}{M_{\lambda_n+1}}$$
$$= r^\omega T_\omega(r).$$

Therefore from the divergence of $\int_1^\infty \log T_\omega(r)dr/r^2$ results the divergence of $\int_1^\infty \log T(r)dr/r^2$, and so under our present hypotheses the class $C_{\{M_n\}}$ is quasi-analytic. Therefore $f(x)\equiv 0$ and the theorem is proved.

## Part II

5. If $f(x)$ belongs to $C^*_{\{M_n\}}$, $G=\limsup_{n\to\infty}(\nu_n/n)$ is always less than or equal to 2. For, the sequence $\{\nu_n\}$ of the orders of the derivatives supposed zero at the origin must contain all the odd integers, since all the odd derivatives are zero there. If $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n}<\infty$ with no further restrictions on $G$, we will prove that $f(x)$ is a trigonometric polynomial

Theorem IV. *If $f(x)$ belongs to $C^*_{\{M_n\}}$ where $\int_1^\infty \log T(r)dr/r^2=\infty$, and if $\limsup_{n\to\infty} |f^{(n)}(0)|^{1/n}<\infty$, then $f(x)$ is a trigonometric polynomial.*

In this case we consider the function $F_1(z)$ defined by (2)

$$F_1(z) = \int_0^{2\pi} f(x)e^{-zx}dx = (1-e^{-2\pi z})\sum_{m=0}^{n-1}\frac{f^{(m)}(0)}{z^{m+1}} + \frac{1}{z^n}\int_0^{2\pi} f^{(n)}(x)e^{-zx}dx.$$

Suppose

$$f(x) = \sum_{m=0}^{\infty} a_m \cos mx.$$

Then

$$F_1(z) = \sum_{m=0}^{\infty} a_m \int_0^{2\pi} e^{-zx}\cos mx\,dx = z(1-e^{-2\pi z})\sum_{m=0}^{\infty}\frac{a_m}{z^2+m^2}\cdot \tag{17}$$

The above termwise integration is valid and the series obtained converges uniformly in any bounded closed region; that is, the remainder $\sum_{m=n}^\infty a_m/(z^2+m^2)$ tends uniformly to zero as $n$ tends to infinity. As in the case of Part I, since $\int_1^\infty \log T(r)\,dr/r^2=\infty$, $F_1(z)$ is given by

$$F_1(z) = (1-e^{-2\pi z})\sum_{m=0}^{\infty}\frac{f^{(m)}(0)}{z^{m+1}}$$

in $|z|>R$, where $R=\limsup_{n\to\infty}|f^{(n)}(0)|^{1/n}<\infty$. But this latter series converges to a holomorphic function in $|z|>R$ and so it follows from (17) that $a_m=0$ for $m>R$. This proves the theorem.

6. In the case in which it is no longer assumed that $\limsup_{n\to\infty}|f^{(n)}(0)|^{1/n}<\infty$, we have the analogue of Theorem III of Part I:

THEOREM V. *Suppose $f(x)$ belongs to $C^*_{\{M_n\}}$ and $f^{(\nu_n)}(0)=0$ $(n\geqq 1)$. If $G=\limsup_{n\to\infty}(\nu_n/n)<2$ and if there exists a positive constant $\omega<2/G-1$ such that $\int_1^\infty \log T_\omega(r)\,dr/r^2=\infty$, then $f(x)$ is identically zero.*

Here again we consider the function $F_1(z)$ and from (2) we have

$$\frac{F_1(z)}{1-e^{-2\pi z}}-\sum_{m=0}^{n-1}\frac{f^{(m)}(0)}{z^{m+1}}=\frac{1}{z^n(1-e^{-2\pi z})}\int_0^{2\pi} f^{(n)}(x)e^{-zx}dx.$$

Then in $\Re(z)\geqq\gamma>0$,

$$\left|\frac{F_1(z)}{1-e^{-2\pi z}}-\sum_{m=0}^{n-1}\frac{f^{(m)}(0)}{z^{m+1}}\right|<\frac{c'k^nM_n}{|z|^n},$$

where $c'$ is a constant.

If we let $z=e^s$ and $\Psi(s)=F_1(e^s)/(1-e^{-2\pi e^s})$, on recalling that $f^{(\nu_n)}(0)=0$ $(n\geqq 1)$, this last inequality becomes

$$\left|\Psi(s)-\sum_{m=1}^{n}f^{(\lambda_m)}(0)e^{-(\lambda_m+1)s}\right|<c'k^{\lambda_n+1}M_{\lambda_n+1}e^{-(\lambda_n+1)\sigma}.$$

But this is precisely inequality (16) with $c$ replaced by $c'$. Therefore in this case also the theorem now follows immediately from Lemma VII.

## BIBLIOGRAPHY

1. Vladimir Bernstein, *Leçons sur les Progrès Récents de la Théorie des Séries de Dirichlet*, Gauthier-Villars, Paris, 1933.

2. T. Carleman, *Les Fonctions Quasi-Analytiques*, Gauthier-Villars, Paris, 1926.

3. A. Denjoy, *Sur les fonctions quasi-analytiques de variable réelle*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 173 (1921), p. 1329.

4. S. Mandelbrojt, *Séries de Fourier et Classes Quasi-Analytiques de Fonctions*, Gauthier-Villars, Paris, 1935.

5. ———, Rice Institute Pamphlet, vol. 29, 1942.

6. ———, *Séries Lacunaires*, Actualités Scientifiques et Industrielles, Exposés sur la théorie des fonctions, 1936.

7. S. Mandelbrojt and J. J. Gergen, *On entire functions defined by a Dirichlet series*, American Journal of Mathematics, vol. 53 (1931), p. 1.

8. A. Ostrowski, *Über Quasianalytische Funktionen und Bestimmtheit Asymptotischer Entwickelungen*, Acta Mathematica, vol. 53 (1930), p. 181.

THE RICE INSTITUTE,
HOUSTON, TEXAS

# A UNIFIED THEORY OF PROJECTIVE SPACES AND FINITE ABELIAN GROUPS

BY

REINHOLD BAER

The similarity between finite-dimensional projective spaces and finite abelian groups has often been noted(1); and thus one may expect that the more general features of these two theories are identical. But the likeness is more than a superficial one; and consequently it is possible to give a unified treatment for spaces and groups.

It may be worthwhile to indicate in a few lines the developments leading up to a scientific situation that made such a joint theory as we are offering here a possibility. The evolution of geometrical thought pertinent to our problem is perhaps best described by two textbooks: Bôcher's *Higher Algebra* which exposed the identity of geometry and the theory of linear equations; and Veblen and Young's *Projective Geometry* whose presentation of the theory broke down the restriction to the two geometries over the real and the complex number field; and enlarged the domain to be considered to the projective geometries over any sort of field, whether finite or infinite, commutative or not. Any further progress had to be a progress in the theory of linear equations; and this was found in the treatment of the theory without using determinants—a concept that had to be thoroughly debunked to make these (and other) extensions possible. This was the starting point for further generalizations, generalizations in a direction that is different from ours—notably the theory of linear equations with infinitely many variables and in particular its geometrical counterpart, J. von Neumann's continuous geometry.

In the theory of finite abelian groups only one generalization was needed. We refer to the extension of this theory by introducing the concept of an abelian group which admits operators from some given ring or some other domain. For this concept makes it possible to consider projective geometry a—rather special—chapter in the theory of abelian groups, since the $n$-dimensional projective space over the (not necessarily commutative) field $F$ of coordinates is nothing but the set of $F$-admissible subgroups of an abelian group of rank $n+1$ over $F$ and the linear forms over this geometry are just the characters of this underlying group. Our problem is now easily stated: to characterize a class of abelian operator groups which comprises both the

Presented to the Society, April 11, 1941; received by the editors February 13, 1941, and, in revised form, September 19, 1941.

(1) See, for example, R. D. Carmichael, *Finite geometries and the theory of groups*, American Journal of Mathematics, vol. 52 (1930), pp. 754–788.

finite abelian groups and the abelian groups over fields as special cases; and to develop a theory of this class of groups which contains finite-dimensional projective geometry and finite commutative group theory as special cases.

It is clear from the preceding remarks that the theory of abelian operator groups may be considered from two rather different points of view, for it is both a special chapter in the theory of groups and a generalization of projective geometry and the theory of linear forms; and according to the point of view preferred, one will write the group composition as "multiplication" or "addition," the operators as "exponents" or "multipliers."

The present investigation has been divided into three parts and this division has been guided by the customary organization of projective geometry. There is first the synthetic theory which deals with the subgroups and their combinations and which does not make any use of the elements of the underlying group, their addition or their multiplication by operators. The analytic theory concerns itself with those facts that either cannot be proved without making use of the group elements and their combinations (or equivalent hypotheses) or which actually involve them in their statement. The third part is devoted to the construction of the underlying abelian operator group to a given partially ordered set meeting certain requirements ( = introduction of coordinates in the terminology of projective geometry) and which may be considered the main contribution of this investigation. Though this part takes an intermediate place between the synthetic and the analytic theory, we had to place it at the end for technical reasons.—A little more detailed account of the contents of these three parts will be found in the introductions prefacing them.

## PART I. THE SYNTHETIC THEORY

The system of admissible subgroups of an abelian operator group is known to form a partially ordered set, containing the sums and cross-cuts of its elements and obeying Dedekind's law, in short, a *Dedekind set*(²). Thus it is natural to make Dedekind sets the framework around which to build the synthetic theory. The atoms of a projective space are its points; and likewise one may consider as the atoms of a finite abelian group its cyclic subgroups of order a power of a prime. Consequently we choose as the atoms of our theory the *cycles*, that is, elements in a Dedekind set whose parts form a finite ordered set. In order to obtain a satisfactory theory including in particular the theory of dimension (or rank), the existence of complementary subspaces and the basis theorem for finite abelian groups, it turns out to be necessary and sufficient to impose the following two conditions upon the

(²) There exists an extensive literature on partially ordered sets, notably the work of G. Birkhoff and O. Ore; for a survey of this theory, see G. Birkhoff, *Lattice Theory*, American Mathematical Society Colloquium Publications, vol. 25, 1940. It should be noted that we use only very elementary parts of this theory, and that we state in full whatever we use.

Dedekind set under consideration. (1) Every element is a sum of cycles. (2) A quotient system is a cycle if, and only if, it contains at most one smallest element not zero (=subcycle of order 1=point).

To obtain a satisfactory geometry one has to assume that lines carry at least three different points. The group-theoretical counterpart is the restriction to primary groups (=groups of order a power of a prime). Thus we have to prove a generalization of the reduction theorem stating that every finite abelian group is the direct sum of its primary components (=partition into relatively prime, primary components); and we show that restriction to primary systems is equivalent to substituting for condition (2) the following condition. (2′) A quotient system is a cycle if, and only if, it contains at most two different smallest elements not zero.

1. **Preliminaries**([3]). The framework for our investigation of the system of admissible subgroups of a (primary) abelian operator group is provided by concepts like "partially ordered set," "lattice," "structure," and so on. Since these systems will be required to satisfy Dedekind's law, we shall term them "Dedekind sets." Such a Dedekind set is a system $D$ of elements connected by the following three relations: the cross-cut, meet, intersection or product $fg$ of the elements $f$, $g$; the join or sum $f+g$ of the elements $f$ and $g$; the relation $f\leqq g$ (in words: $f$ is a part of or contained in $g$). The rules by means of which the first two operations may be reduced to the third one, and conversely, may be stated as follows:

$fg$ is the greatest element contained in both $f$ and $g$.

$f+g$ is the smallest element containing both $f$ and $g$.

$f\leqq g$; $f=fg$ and $f+g=g$ are equivalent assertions.

Let us add finally that the relation: $f\leqq g$ is reflexive and transitive; and that $f<g$ signifies: $f\neq g$ but $f\leqq g$.

To these elementary rules we add the existence of a null-element 0 and we impose the main requirement, namely

DEDEKIND'S LAW. *If $f$, $g$, $h$ are three elements in $D$, and if $f\leqq g$, then $f+gh=(f+h)g$.*

This is a partial substitute for the distributive law; and we are going to derive from it some further useful formulas.

LEMMA I.1.1. *If the four elements $a$, $b$, $c$, $d$ in the Dedekind set $D$ satisfy $(a+b)(c+d)=0$, then*

$$(a+c)(b+d)=ab+cd=(a+d)(b+c).$$

**Proof.** It suffices to prove the first of these equations; and this may be done as follows:

([3]) In this section we collect a number of elementary facts from the theory of partially ordered sets in the form best suited to our purposes (see Footnote 2).

$$\begin{aligned}(a + c)(b + d) &= (a + c)(a + b + c)(b + d) = (a + c)(b + d)(a + b + c)\\ &= (a + c)(b + d(d + c)(a + b + c))\\ &= (a + c)(b + d(c + (d + c)(a + b)))\\ &= (a + c)(b + dc) = dc + b(a + c) = dc + b(a + b)(a + c)\\ &= dc + b(a + c(a + b)) = dc + ab.\end{aligned}$$

LEMMA I.1.2. *If $a$, $b$, $c$ are three elements in $D$ such that $ab=(a+b)c=0$, then $a(b+c)=b(c+a)=0$.*

For $a(b+c)=a(a+b)(b+c)=a(b+c(a+b))=ab=0$.

The elements $x_1, \cdots, x_n$ are said to be *independent*, if $x_i\sum_{j\neq i}x_j=0$ for $i=1, \cdots, n$; and it is readily inferred(4) from Lemma I.1.1 that the elements $x_1, \cdots, x_n$ *are independent if, and only if,* $0=x_i\sum_{j<i}x_j$ *for* $i=2, \cdots, n$.

If the elements $x_1, \cdots, x_n$ are independent, then $\sum_{i=1}^n x_i=s$ is the *direct sum* of the $x_i$ and every summand $x_i$ is a *direct summand* of $s$. The following statements are easily verified. If $x$ is the direct sum of the $x_i$, and if $x_i$ is the direct sum of the elements $x_{ij}$, then $x$ is the direct sum of the $x_{ij}$. If $s$ is the direct sum of $t$ and $u$, and if $v$ is an element between $u$ and $s$, then $v$ is the direct sum of $u$ and $vt$; so that $a$ is a direct summand of $b$, if $a$ is a direct summand of $c$ and if $a\leqq b\leqq c$.

If $u\leqq v$, then the set $v/u$ of all the elements $x$ in $D$ which satisfy $u\leqq x\leqq v$ is a Dedekind set (exactly as $D$) and $u$ is the null-element of $v/u$. If in particular $u=0$, then $v/0$ is the set of all the parts of $v$; and it will be possible to write $v$ instead of $v/0$ without causing confusion.

Any biunivoque and monotonically increasing correspondence, mapping the elements of the Dedekind set $D$ upon the elements of $D'$ is termed an *isomorphism* or a *projectivity*(5) of $D$ upon $D'$. If in particular $u$ and $v$ are two elements in $D$, then *an isomorphism of* $(u+v)/u$ *upon* $v/(uv)$ is defined by mapping $x$ in $(u+v)/u$ upon $xv$ in $v/(uv)$.

2. **Cycles and their orders.** If $u$ and $v$ are elements in a partially ordered set (in particular in a Dedekind set), then $u$ is said to be a *cycle modulo* $v$, if $v\leqq u$, if there exists only a finite number of elements between $v$ and $u$, and if of two elements between $u$ and $v$ one is always part of the other, that is, if $u/v$ is a finite ordered set. If $u$ is a cycle modulo $v$, then we denote by $n(u/v)$ the number of elements between $u$ and $v$ which are different from $v$ and term this number the *order* of the cycle $u/v$. Instead of $n(u/0)$ we write $n(u)$ and we say that $u$ is a cycle instead of saying a cycle modulo 0. If the

(4) Cf. K. Menger, Annals of Mathematics, (2), vol. 37 (1936), pp. 456–482.

(5) As long as only partially ordered sets are discussed, we prefer the term "isomorphism" as the more appropriate one. But as soon as we have to connect the "isomorphisms" of partially ordered sets (of subgroups of a group) with the "isomorphisms" of groups, we shall use the term "projectivity" in order to avoid confusion.

cycle $z$ is part of the element $e$, then we express this fact by saying that $z$ is a *subcycle of e*.

THEOREM I.2.1. *If $z$ is a subcycle of a sum of cycles whose orders do not exceed $m$, then the order of $z$ does not exceed $m$ either.*

**Proof.** It is a well known fact[6] that every part of a sum of a finite number of cycles of order 1 (of points in projective geometry) is itself a sum of a finite number of cycles of order 1; and from this fact one readily infers our theorem in the special case $m=1$.

We now proceed by induction with regard to $m$, assuming the theorem to be true for $m$ and proving it for $m+1$. Let $s$ be the sum of the cycles $z_i$ of order not exceeding $m+1$; and suppose that $z$ is a subcycle of $s$. If $n(z_i)=m+1$, then denote by $y_i$ the uniquely determined subcycle of order $m$ of $z_i$; and if $n(z_i)\leqq m$, then put $z_i=y_i$. If $t$ is the sum of the $y_i$, then it follows from the induction hypothesis that the subcycle $tz$ of $t$ has an order not exceeding $m$. We note furthermore that $s/t$ is the sum of the cycles $(t+z_i)/t$ whose orders do not exceed 1, since they are isomorphic to $z_i/(z_it)$ and since $y_i\leqq z_it$. Thus it follows from the special case $m=1$ that the subcycle $(z+t)/t$ of $s/t$ is of an order not exceeding 1 so that $n(z/(zt))\leqq 1$. Since we already pointed out that $n(zt)\leqq m$, we find now that $n(z)=n(zt)+n(z/(zt))\leqq m+1$, as was to be shown.

THEOREM I.2.2. *If $z$ is a subcycle of the direct sum $s$ of the cycles $z_i$, if $n$ is a positive integer, if $y_i$ is the uniquely determined subcycle of $z_i$ whose order is exactly the minimum of the numbers $n$ and $n(z_i)$, and if $y$ is the sum of the cycles $y_i$, then $n(z)\leqq n$ is a necessary and sufficient condition for $z\leqq y$.*

REMARK. It is easy to give examples showing that the independence of the cycles $z_i$ is indispensable for the validity of this theorem.

**Proof.** It is a consequence of Theorem I.2.1 that the orders of the subcycles of $y$ do not exceed $n$, since $n(y_i)\leqq n$.—Thus assume conversely that $n(z)\leqq n$. If $k$ is the number of cycles $z_i$, then we put $s(i)=\sum_{j=1}^{i}y_j+\sum_{j=i+2}^{k}z_j$ and $t(i)=s(i)+z_{i+1}$ so that in particular $t(0)=s$ and $t(k)=y$. Since $z$ is a subcycle of $t(0)$ we are going to prove by complete induction with regard to $i$ that $z$ is a subcycle of each $t(i)$. Thus we assume that $z\leqq t(i)$ and we have to prove that $z\leqq t(i+1)$. From the induction hypothesis it follows that $s(i)\leqq s(i)+z\leqq t(i)=s(i)+z_{i+1}$. Since $(s(i)+z)/s(i)$ and $z/(zs(i))$ are isomorphic cycles, and since the order of $z$ does not exceed $n$, it follows that $n((s(i)+z)/s(i))\leqq n$. Since $s$ is the direct sum of the cycles $z_j$, it follows that $0=s(i)z_{i+1}$ so that $z_{i+1}$ and $t(i)/s(i)$ are isomorphic cycles. Consequently $s(i)+z\leqq s(i)+y_{i+1}=t(i+1)$; and this completes the proof.

We note without proof the important fact that *the maximum and the minimum conditions are satisfied by the parts of a sum of a finite number of*

(6) Cf. Menger, loc. cit.

*cycles*; it is an obvious consequence of the fact(7) that the maximum and minimum conditions are satisfied by the parts of $a+b$, if they are satisfied by the parts of $a$ and of $b$.

3. **Direct decompositions.** The part $v$ of the element $w$ (in the Dedekind set $D$) is said to be *closed in*(8) $w$, if to every cycle $z \leqq w$ such that $zv \neq 0$ there exists a cycle of order $n(z)$ between $zv$ and $v$.

THEOREM I.3.1. *Every direct summand of $w$ is closed in $w$.*

**Proof.** If $w$ is the direct sum of $u$ and $v$, if $z$ is a subcycle of $w$ such that $vz \neq 0$, then $c = v(u+z)$ is between $zv$ and $v$. Thus $cu=0$ and $zu=0$ are consequences of the fact that the only subcycle of order 1 of the cycle $z$ is in $v$. Hence $c = c/(cu)$ is isomorphic to $(c+u)/u = (v(u+z)+u)/u = ((v+u)(u+z))/u = (w(u+z))/u = (u+z)/u$ as follows from Dedekind's law; and $(u+z)/u$ being isomorphic to $z/(uz) = z$, it follows that $c$ and $z$ are isomorphic so that they are cycles of equal order.

THEOREM I.3.2. *If $z$ is a subcycle of maximum order of the direct sum $s$ of the cycles $c(1), \cdots, c(k)$, then there exists an $i$ such that $s$ is the direct sum of $z$ and of the cycles $c(j)$ for $j \neq i$.*

**Proof.** We may assume that $n(c(i)) = n(z)$ if, and only if, $1 \leqq i \leqq h$; and we note that $0 < h$ by Theorem I.2.1. If $v(i) = \sum_{j \neq i} c(j)$, then $\prod_{i=1}^{h} v(i) = \sum_{h<j} c(j) = v$. Since $v$ is a direct summand of $s$, $v$ is closed in $s$ by Theorem I.3.1; and since the maximum order of the subcycles of $v$ is smaller than $n(z)$—by Theorem I.2.1—$vz=0$. Thus there exists at least one $i$ between 1 and $h$ such that the subcycle $z^*$ of order 1 of $z$ is not contained in $v(i)$; since thus $zv(i)=0$, and since $s$ is the direct sum of $v(i)$ and of the cycle $c(i)$ of order $n(z)$, it follows finally that $s$ is the direct sum of $z$ and of $v(i)$.

COROLLARY I.3.3. *Every part of $w$ is a direct sum of cycles if, and only if, the following conditions are satisfied.*

(i) *Every part of $w$ is a sum of cycles.*

(ii) *If $z$ is a subcycle of maximum order of the part $v$ of $w$, then $z$ is a direct summand of $v$.*

The necessity is an immediate consequence of Theorem I.3.2, the sufficiency may be proved inductively, since every part $v \neq 0$ of $w$ is the direct sum of a subcycle of maximum order (in $v$) and of some smaller element, and since (i) implies the minimum condition for the parts of $w$.

COROLLARY I.3.4. *If the element $s$ is both the direct sum of the cycles $c(i)$*

(7) Cf. Birkhoff, loc. cit.

(8) This concept has been introduced by H. Prüfer into the theory of primary abelian groups (under the name "Servanzuntergruppe"); cf. H. Prüfer, Mathematische Zeitschrift, vol. 17 (1923), pp. 35–61.

***and the direct sum of the cycles $d(j)$, then the number of cycles $c(i)$ of order $n$ is the same as the number of cycles $d(j)$ of order $n$.***

For if $d(1)$ is a cycle of maximum order in $s$, then it follows from Theorem I.3.2 that $s$ is the direct sum of $d(1)$ and of the $c(i)$ for $i \neq k$. Thus $d(1)$ and $c(k)$ are of the same order and $\sum_{1<i} d(i)$ and $\sum_{j\neq k} c(j)$ are isomorphic; and now the statement may be proved by induction.

The element $w$ ***splits***, if every part of $w$ is a direct sum of (a finite number of) cycles, and if every closed part of any element $v \leqq w$ is a direct summand of $v$. The following characterization of splitting elements will be needed in the proof of the main theorem of this section.

THEOREM I.3.5. *The element $w$ splits if, and only if, it satisfies the following conditions.*

(i) *Every part of $w$ is a sum of cycles.*

(ii) *If $t < s \leqq w$, if $t$ is a subcycle of maximum order of $s$, and if $s/t$ is a cycle, then $s$ contains a cycle of order* 1 *which is not part of $t$; and if $p$ is a subcycle of order* 1 *not part of $t$, then $s$ is the direct sum of $t$ and of a cycle containing $p$.*

**Proof.** If $w$ splits, and if $s$ and $t$ satisfy the hypotheses of (ii), then $t$ is closed in $s$, therefore a direct summand of $s$ so that $s$ is the direct sum of $t$ and of some cycle $z$. Thus there exist subcycles of order 1 of $s$ which are not in $t$. If $p$ is some such cycle of order 1, then denote by $c$ a cycle of greatest order between $p$ and $v$. Then $c$ is closed in $v$ and therefore a direct summand of $v$. Since $n(z) \leqq n(t)$, it follows now from Corollary I.3.4 that $n(c) = n(z)$; and $tc = 0$ implies now that $v$ is the direct sum of $c$ and $t$. Thus splitting elements satisfy (i) and (ii).

For the sufficiency proof it will be convenient to say that the part $r$ of $s$ is ***weakly closed in** $s$*, if subcycles of order 1 of $r$ which are contained in subcycles of order $n$ of $s$ are contained in subcycles of the same order $n$ of $r$.

Suppose now that $w$ satisfies (i) and (ii), that $r < v \leqq w$ and that $r$ is weakly closed in $v$. Since $r$ and $v$ are sums of cycles, there exists a subcycle $x$ of smallest order of $v$ which is not part of $r$. If $x^*$ is the subcycle of order 1 of $x$, then $x^* \leqq r$ would imply the existence of a cycle $y$ of order $n(x)$ between $x^*$ and $r$. It follows from Theorem I.2.1 that $n(y)$ is the maximum order of the subcycles of $x+y$; and thus it follows from (ii) that $y$ is a direct summand of $x+y$ which proves the existence of a cycle of order smaller than $n(x)$ which is part of $v$ but not of $r$. This contradiction shows that $xr = 0$. Thus there exists a subcycle $z$ of $v$ such that $zr = 0$ and such that the order of $z$ is as big as possible. Since $z \neq 0$, this implies $r < r+z$. Suppose now that $p$ is a subcycle of order 1 of $r+z$, that $b$ is a cycle between $p$ and $v$. If $p \leqq r$, then there exists a cycle of order $n(b)$ between $p$ and $r$. If the inequality $p \leqq r$ does not hold, then $br = 0$ so that $n(b) \leqq n(z)$. Thus we may assume that $pz = 0$ in order to

prove that $r+z$ is weakly closed in $v$. Then $q=r(p+z)$ is a cycle of order 1, since $p$ and $(p+z)/z=(q+z)/z$ are isomorphic. Since $qz=0$, it follows from (ii) that $b+z$ is the direct sum of $z$ and of a cycle $d$ containing $q$. Since $bz=0$, $n(b)=n(d)$. Since $q\leqq r$, and since $r$ is weakly closed in $v$, there exists a cycle $e$ of order $n(d)$ between $q$ and $r$. Since $p\leqq q+z\leqq e+z$, there exists by (ii) a cycle $f$ containing $p$ such that $e+z$ is the direct sum of $f$ and $z$. It follows from Corollary I.3.4 that $n(f)=n(d)=n(b)$; and thus we have finally shown that $r+z$ is weakly closed in $v$. Since by (i) the maximum condition is satisfied by the parts of $w$, it follows now by induction that $v$ is the direct sum of $r$ and of some cycles. Thus we have shown that (i), (ii) imply the splitting of $w$ and imply that *every weakly closed part is a direct summand* (and therefore closed).

If the element $v$ is part of the element $u$, then $u$ splits modulo $v$, whenever $u$ splits in the Dedekind set $u/v$ (which consists of all the elements between $v$ and $u$ and whose null is $v$). The element $w$ *splits completely*, if every part $u$ of $w$ splits modulo each of its parts $v$. We mention the important and well known fact that both finite abelian groups[9] and finite-dimensional projective geometries[10] split completely.

THEOREM I.3.6. *The element $w$ splits completely if, and only if, it satisfies the following conditions.*

(i) *Every part of $w$ is a sum of cycles.*

(ii) *If $r\leqq s\leqq w$, and if $s/r$ contains at most one subcycle of order* 1, *then $s/r$ is a cycle.*

**Proof.** The necessity of these conditions is an immediate consequence of the fact that subcycles of maximum order are closed, and that $s/r$ is a sum of cycles, if $s$ is a sum of cycles.

Before proving the sufficiency of (i) and (ii) we prove the following helpful lemma.

(I.3.6.1) *If $s$ is a sum of cycles, if $t<s$ and if $s/t$ is a cycle then there exists a cycle $z$ such that $s=t+z$.*

For there exists between $t$ and $s$ one and only one element $r$ such that $s/r$ is a cycle of order 1. Since $r<s$, not every subcycle of $s$ is part of $r$. If $z$ is a subcycle of $s$, though not of $r$, then the inequality $t+z\leqq r$ does not hold so that $t+z=s$, since $s/t$ is a cycle.

Suppose now that (i) and (ii) are satisfied by $w$, that $t<s\leqq w$, that $t$ is a subcycle of maximum order of $s$ and that $s/t$ is a cycle. Since $s$ is therefore not a cycle, it follows from (ii) that $s$ contains at least two different subcycles of order 1, one of which is certainly not part of $t$. Suppose now that $p$ is a sub-

[9] Cf. Prüfer, loc. cit.

[10] The central importance of this fact for projective geometry has been stressed by Menger, loc. cit., and by G. Birkhoff, Annals of Mathematics, (2), vol. 36 (1935), pp. 743–748.

cycle of order 1 of $s$ and that the inequality $p \leqq t$ does not hold or $pt=0$. Denote by $b$ a cycle of greatest order between $p$ and $s$. If $g$ is an element between $b$ and $s$ such that $g/b$ is a cycle of order 1, then $g$ is no cycle so that $g$ contains by (ii) a subcycle $q$ of order 1 different from $p$. Since it follows from (I.3.6.1) that $s$ is the sum of $t$ and of some other cycle, it follows that the sum $s^*$ of all the subcycles of order 1 of $s$ is the sum of any two of its (different) subcycles (of order 1) so that $s^*=p+q$ or $g=b+s^*$. Thus $s/b$ contains one and only one subcycle of order 1; and (ii) implies consequently that $s/b$ is a cycle. Hence it follows from (I.3.6.1) that $s$ is the sum of $b$ and of some cycle so that $n(s/b) \leqq n(t)$. Since $tb=0$, it follows now that $s$ is the direct sum of $t$ and of the cycle $b$ containing $p$. Thus we showed that the conditions of Theorem I.3.5 are satisfied by $w$, if $w$ satisfies (i) and (ii); that is, if $w$ satisfies our conditions (i) and (ii), then $w$ splits. But if $w$ satisfies conditions (i) and (ii), and if $u \leqq v \leqq w$, then $v/u$ satisfies these conditions so that $v/u$ splits too. Thus $w$ splits completely, if it satisfies the conditions (i) and (ii).

The elements $u$ and $v$ are termed *relatively prime*, if $u$ and $v$ are not both 0, if $uv=0$, and if $x \leqq u+v$ implies $x=xu+xv$. The decomposition of a finite abelian group into its primary components is an example of a decomposition into a sum of relatively prime elements. The elements $u$ and $v$ are said to be relatively prime modulo their common part $t$, if they are relatively prime elements in the Dedekind set $(u+v)/t$. Finally we say that the element $w$ is *primary*, if there does not exist any triplet of elements $u$, $v$, $t$ such that $t \leqq uv \leqq u+v \leqq w$ and such that $u$ and $v$ are relatively prime modulo $t$. The system of subgroups of an abelian group of order a power of a prime furnishes an example of a primary system; and projective geometries whose lines carry at least three points are primary too.

If the sum of the two cycles $p$ and $q$ of order 1 contains just these two cycles and no further cycles (of order 1), then $p$ and $q$ are relatively prime.—If $u$ and $v$ are relatively prime, and if $u$ and $v$ are sums of cycles, then $u$ contains a cycle $p$ of order 1, $v$ contains a cycle $q$ of order 1, and $p+q$ contains just these two cycles and no further ones. Combining these remarks with Theorem I.3.6 we obtain the following fundamental theorem.

THEOREM I.3.7. *The element $w$ splits completely and is primary if, and only if, it satisfies the following conditions.*

(i) *Every part of $w$ is a sum of cycles.*

(ii) *If $r \leqq s \leqq w$, and if $s/r$ contains at most two different subcycles of order* 1, *then $s/r$ is a cycle.*

An $n$-dimensional projective geometry whose lines carry at least three points possesses systems of $n+1$ points no $n$ of which are on a hyperplane. To generalize this property which will be of importance in the future we say that the elements $v(i)$ form a *partial sum* of the elements $u(i)$, if $v(i) \leqq u(i)$ for $i=1, \cdots, n$. If we have $v(i) < u(i)$ for at least one $i$, then the $v(i)$ form a

*proper partial sum* of the $u(i)$. If the $u(i)$ are independent, and if the $v(i)$ form a partial sum of the $u(i)$, then they form a proper partial sum if, and only if, $\sum_{i=1}^{n} v(i) < \sum_{i=1}^{n} u(i)$.

LEMMA I.3.8. *If the primary element $s$ is the direct sum of the cycles $c(i)$, if the parts of $s$ are sums of cycles, then there exists a subcycle of $s$ which is not contained in any proper partial sum of the $c(i)$; if the subcycle $z$ of $s$ is not contained in any proper partial sum of the $c(i)$, then $z$ is a subcycle of maximum order of $s$ and $s$ is the direct sum of $z$ and of the $c(i)$ for $i \neq k$, provided $c(k)$ is of maximum order too.*

**Proof.** If we denote by $c(i)'$ the uniquely determined subcycle of $c(i)$ such that $c(i)/c(i)'$ is a cycle of order 1 and by $s'$ the sum of the $c(i)'$, then $s/s'$ is an $(n-1)$-dimensional projective geometry (a direct sum of the cycles $c(i)/c(i)'$ of order 1). Hence there exists in $s/s'$ a cycle $c$ of order 1 which is not part of any proper partial sum of the $c(i)/c(i)'$. But $c$ may be seen to be the sum of $s'$ and of a cycle $z$ which meets the requirements.—The second statement is obvious.

4. **Partition into relatively prime, primary components.** The elements $u_1, \cdots, u_k$ *constitute a partition of their sum* $s$, if $s$ is the direct sum of the $u_i$, and if $u_i$ and $\sum_{j\neq i} u_j$ are relatively prime for every $j$. Note that the primary components of a finite abelian group constitute a partition of this group. If $w$ is a completely splitting element in a Dedekind set, then $w$ is the direct sum of cycles and therefore of primary elements; but there exist examples of completely splitting elements which do not admit of a partition into relatively prime, primary elements.

THEOREM I.4.1. (a) *The element $w$ in the Dedekind set $D$ admits of at most one partition into (relatively prime) primary elements.* (b) *The completely splitting element $w$ admits of a partition into (relatively prime) primary elements if, and only if, any two subcycles with non-primary sum are relatively prime.*

**Proof.** Assume first that the elements $u_i$ as well as the elements $v_i$ constitute a partition of the element $w$. Then the elements $u_i v_j \neq 0$ for $j = 1, \cdots, h$ constitute a partition of $u_i$ so that the elements $u_i v_j \neq 0$ constitute a partition of $w$. Statement (a) is now a consequence of the fact that primary elements do not admit of partitions into (more than one) relatively prime element.

Suppose now that the primary elements $p_1, \cdots, p_k$ constitute a partition of the completely splitting element $w$. If $z$ is a subcycle not 0 of $w$, $z^*$ the uniquely determined subcycle of order 1 of $z$, then $z = \sum_{i=1}^{k} z p_i$, $z^* = \sum_{i=1}^{k} z^* p_i$. Consequently there exists a subscript $j$ such that $z^* p_j = z^*$, $z^* p_i = 0$ for $i \neq j$; and this implies $z \leq p_j$, $z p_i = 0$ for $i \neq j$ so that *every subcycle of $w$ is contained in one and only one component $p_i$.*

If $u$ and $v$ are two subcycles of $w$, then they are either contained in the same component $p_i$—in which case their sum is primary—or else they are in

different components and then they are relatively prime proving the necessity of the condition of (b).

Suppose now that the condition of (b) be satisfied by the subcycles of the completely splitting element $w$. If the part $t$ of $w$ is not primary, then there exist elements $x$, $y$ such that $x<y\leqq t$ and such that $y/x$ consists of exactly four elements, namely $x$, $y$ and two cycles of order 1. Since $w$ and therefore $t$ splits completely, this implies the existence of two subcycles $u$, $v$ not 0 of $t$ such that $uv=0$ and such that $u+v$ is not primary. Hence it follows from the hypothesis that $u$ and $v$ are relatively prime.—Suppose now that $s<t$, that $t/s$ is a cycle of order 1, and that $s$ is primary. If we denote by $c'$ the uniquely determined subcycle of order $n(c)-1$ of the cycle $c\neq 0$, then $c'\leqq s$ for every subcycle $c\neq 0$ of $t$ so that in particular $u'+v'\leqq s$. But the inequality $u+v\leqq s$ does not hold since $u+v$ is not primary, though $s$ is primary. Thus not both cycles $u$ and $v$ are subcycles of $s$. Since $t/s=(s+u+v)/s$ is a cycle of order 1 and is isomorphic to $(u+v)/s(u+v)=(u+v)/(su+sv)$—as $u$ and $v$ are relatively prime—it follows now that at least one of the cycles $u$ and $v$ is part of $s$. Thus we assume that $u\leqq s$, and that the inequality $v\leqq s$ does not hold. Since $u+v'\leqq s$, $u+v'$ is primary; and since $u$ and $v$ are relatively prime, this implies $v'=0$ so that $v$ is a cycle of order 1. If now $c$ is a subcycle of order 1 of $s$ which is different from the subcycle $u^*$ of order 1 of $u$, then $u^*+c$ is the direct sum of two cycles of order 1 and is primary as a part of $s$. Hence there exists a subcycle $z$ of order 1 of $u^*+c$ which is different from $c$ and $u^*$ so that $c+u^*=u^*+z=z+c$. Then $c+v=(c+v)/z(c+v)$ is isomorphic to $(z+c+v)/z=(z+u^*+v)/z$ and this is isomorphic to $(u^*+v)/z(u^*+v)=u^*+v$ so that $c$ and $v$ are relatively prime too. If $q$ is any subcycle not 0 of $s$, $q^*$ its subcycle of order 1, then $q^*$ and $v$ are relatively prime; hence $q+v$ is not primary and it follows from the hypothesis that $q$ and $v$ are relatively prime. Since every part of $s$ is a direct sum of cycles, it follows now that $s$ and $v$ are relatively prime. Thus $s$ and the cycle $v$ of order 1 constitute a partition of $t=s+v$ so that every subcycle of $t$ is either part of $s$ or of $v$; and this shows in particular that $v$ is the only subcycle $r$ of $t$ such that $t=s+r$.

Since the maximum condition is satisfied by the parts of $w$, there exists some greatest primary part $p$ of $w$. Certainly $p\neq 0$, if $w\neq 0$, since every subcycle of $w$ is primary. If $p=w$, then $w$ is primary so that we may assume that $0<p<w$. If $z$ is any subcycle of $w$, then either $z\leqq p$ or $p+z$ is not primary. In the latter case $(p+z)/p$ is a cycle different from 0 and there exists a uniquely determined subcycle $c$ of $z$ such that $(p+c)/p$ is a cycle of order 1. From what we have shown in the preceding paragraph of the proof it follows that $c$ is a cycle of order 1 such that $p$ and $c$ are relatively prime. If $x$ is any subcycle of $p$, then $x$ and $c$ are relatively prime; since $x+c\leqq x+z$, it follows that $x+z$ is not primary; and hence it follows from the hypothesis that $x$ and $z$ are relatively prime; and this shows that $p$ and $z$ are relatively prime.

Thus every subcycle of $w$ which is not part of $p$ is relatively prime to $p$.—Denote now by $q$ the sum of all the subcycles of $w$ which are not part of $p$. Then $q=z_1+\cdots+z_h$ where each of the $z_i$ is relatively prime to $p$. If $c$ is a subcycle of order 1 of $p$, then $c$ and $z_1$ are relatively prime. Since $cz_1=0$, we may assume that $cq_{i-1}=0$ for $q_{i-1}=\sum_{j=1}^{i-1}z_j$. If $c(q_{i-1}+z_i)\neq 0$, then $c\leqq q_{i-1}+z_i$. Hence $c+z_i=(c+z_i)(q_{i-1}+z_i)=z_i+q_{i-1}(c+z_i)=z_i$, since $c$ and $z_i$ are relatively prime, and since therefore $q_{i-1}(c+z_i)=q_{i-1}c+q_{i-1}z_i=q_{i-1}z_i$ by the induction hypothesis; but this would imply $c\leqq z_i$, an impossibility which proves $cq=0$. This implies $zq=0$ for every subcycle $z$ of $p$; and hence every subcycle not 0 of $w$ is part of one and only one of the two elements $p$ and $q$. Since the parts of $w$ are sums of cycles, this implies that $p$ and $q$ are relatively prime; and from the validity of maximum and minimum condition for the parts of $w$ one now readily infers that the greatest primary parts of $w$ constitute a partition of $w$ into relatively prime, primary elements.

5. **Sums and products of infinite sets.** It is well known that a great part of the theory of finite abelian groups, in particular the basis theorem, holds true for abelian groups the orders of whose elements are bounded. Cross-cuts of any number of subgroups and sums of any number of subgroups are subgroups too; and we propose to give in this section a short analysis of the pertinent concepts.

If $S$ is a set of elements in the Dedekind set $D$, then the element $p$ in $D$ is a product of $S$, if $p$ is a greatest element contained in all the elements in $S$; and the element $s$ in $D$ is a sum of $S$, if $s$ is a smallest element containing every element in $S$. It is readily verified that there exists at most one product and at most one sum of $S$.

The set $S$ of elements not 0 in $D$ is independent, if there exists for every element $t$ in $S$ the sum $S(t)$ of the elements different from $t$ in $S$, and if $tS(t)=0$ for every $t$ in $S$.

THEOREM I.5.1[11]. *Suppose that the element $w$ satisfies the following conditions.*

(i) *If the nonvacuous set $T$ of subcycles of $w$ contains every subcycle of any finite sum of cycles in $T$, then there exists one and only one part $s(T)$ of $w$ such that $T$ is the set of all the subcycles of $s(T)$.*

(ii) *The orders of the subcycles of $w$ are bounded.*

*Then every part of $w$ is the direct sum of each of its closed parts and of a finite or infinite number of cycles if (and only if) every finite sum of subcycles of $w$ splits.*

REMARK. Condition (i) is satisfied in every abelian group without elements of infinite order and is satisfied in the primary abelian operator groups too (see below).—That condition (ii) is indispensable for the validity of the theorem is well known.

(11) H. Prüfer proved this theorem for primary abelian groups.

**Proof.** We note first that on account of condition (i) sums and products of any number of parts of $w$ exist, that furthermore the subcycle $c$ of $w$ is part of the sum of the set $S$ of parts of $w$ if (and only if) there exists a finite number of elements in $S$ whose sum contains $c$, and that finally the set $S$ of parts not 0 of $w$ is independent if, and only if, every finite subset of $S$ is independent. We recall furthermore that—as in the proof of Theorem I.3.5—the element $u$ is termed weakly closed in the element $v$, if $u \leqq v$, and if there exists a cycle of order $n$ between $p$ and $u$ whenever $p$ is a subcycle of order 1 of $u$ which is contained in a subcycle of order $n$ of $v$.

Suppose now that $u < v \leqq w$ and that $u$ is weakly closed in $v$. Then there exists a subcycle of $v$ which is not part of $u$ and one verifies—exactly as in the proof of Theorem I.3.5—that there exist subcycles not 0 of $v$ which are independent of $u$. Hence it follows from condition (ii) that there exists a subcycle $z$ of $v$ such that $zu = 0$ and such that the order $n(z)$ is as big as possible. To show that the (direct) sum $u+z$ is weakly closed we need consider only subcycles $p$ of order 1 of $u+z$ which satisfy: $pu = pz = 0$. Suppose that $b$ is some cycle between $p$ and $v$. Then $n(b) \leqq n(z)$ and $bz = 0$. That $q = u(p+z)$ is a cycle of order 1 is verified as in the proof of Theorem I.3.5. Thus it follows from Theorem I.3.5 and the fact that the direct sum $b+z$ of the two cycles $b$ and $z$ splits, that there exists a cycle of order $n(b)$ between $q$ and $b+z \leqq v$. Since $q \leqq u$, and since $u$ is weakly closed in $v$, there exists a cycle $d$ of order $n(b)$ between $q$ and $u$. Since $qz = 0$, and since the direct sum $d+z$ of the two cycles $d$ and $z$ splits by hypothesis, it follows from Theorem I.3.5 that there exists a cycle of order $n(d) = n(b)$ between $p$ and $d+z \leqq u+z$; and thus it has been shown that $u+z$ is weakly closed in $v$ too.

If the part $r$ of the element $v \leqq w$ is weakly closed in $v$, then denote by $R$ the set of all the elements $s$ between $r$ and $v$ such that $s$ is weakly closed in $v$ and such that $s$ is the direct sum of $r$ and of a finite or infinite number of cycles. If $x$ and $y$ are two elements in $R$, then $x$ is said to be better than $y$, if $x$ is the direct sum of $y$ and of a finite or infinite number of cycles. From the remarks in the first paragraph of this proof it may be inferred that there exists a best element in $R$; and it is an immediate consequence of the results of the second paragraph of the proof that $v$ itself is the only best element in $R$ so that $v$ is the direct sum of $u$ and of a finite or infinite number of cycles.

If the element $w$ satisfies condition (i) of Theorem I.5.1, then it follows from Theorem I.2.1 that there exists one and only one part $w_n$ of $w$ such that the set of subcycles of $w_n$ is just the set of subcycles of $w$ with an order not exceeding $n$. Every finite sum of subcycles of $w$ is contained in almost every $w_n$; and Theorem I.5.1 may be applied on the $w_n$.

## Part II. The analytic theory

An abelian operator group may be termed a primary abelian operator group, if the system of its admissible subgroups meets the requirements

imposed in the first part. Not only do abelian groups of prime power order and the abelian operator groups underlying projective geometry belong into this class, but it is even possible to develop a theory of primary abelian operator groups which is fully comparable to both projective geometry and the theory of finite abelian groups. For example, the duality between group and character group may be proved, a fact that specializes, in the case of projective geometry, to the duality between the point space and the hyperplane space and which thus contains the theory of systems of linear equations. Further examples are extensions of the fundamental theorem of projectivity and of the theorem of Pappus.

For our purposes it does not suffice to characterize the primary abelian operator groups as operator groups with specific properties. We have to solve the problem of determining those sets of subgroups of an abelian group which are the systems of all the admissible subgroups of a primary abelian operator group. A set $L$ of subgroups of the abelian group $G$ may be proved to be the system of all the admissible subgroups of $G$ for a suitable set of operators, if $L$ satisfies the following conditions: (a) $L$ contains sums and crosscuts of its subsets. (b) If the subgroup $Z$ in $L$ is the smallest subgroup in $L$ containing a given element $z$, then the subgroups in $L$ that are part of $Z$ form a finite ordered set. (c) If a subgroup $Z$ in $L$ contains just $n$ subgroups in $L$ and if the subgroups in $L$ that are part of $Z$ form an ordered set, then there exist at least three independent subgroups of this kind in $L$. Under the same hypotheses we may prove that every projectivity of $L$ is induced by an isomorphism of the underlying group $G$. It seems noteworthy that both these theorems are obtained as special cases from one and the same construction.

1. **Construction of endomorphisms and isomorphisms.** The composition of the elements in commutative groups $G$ will be written as addition: $x+y$. A *linear transformation*(12) of $G$ into the commutative group $H$ is a function $f$ which maps every element $g$ in $G$ upon a uniquely determined element $g^f$ in $H$ such that $(g\pm h)^f=g^f\pm h^f$. Linear transformations of $G$ into $G$ are termed *endomorphisms*(13) *of* $G$; and these shall be written usually as multipliers, that is, the endomorphism $f$ of $G$ maps the element $g$ in $G$ upon the element $gf$ in $G$.

If $E$ is a set of endomorphisms of $G$, $S$ is a subset of $G$, then $SE$ is the set of all the elements $se$ for $s$ in $S$, $e$ in $E$; and the subset of $G$ is *E-admissible*, if $SE\leqq S$. The system of all the $E$-admissible subgroups of $G$ shall be denoted by $D(G; E)$. It is one of the objects of this section to determine all the systems $D(G; E)$ meeting certain requirements.

If $L$ is a set of subgroups of $G$, then the endomorphism $e$ of $G$ is *L-admissible*, if $Se\leqq S$ for every subgroup $S$ in $L$; and the set $K(G; L)$ of all the $L$-admissible endomorphisms of $G$ is a ring, provided addition and multiplica-

(12) Often termed "homomorphism."

(13) Often called "auto-homomorphism," "(proper or improper) automorphism," and so on.

tion are defined as customary. If in particular $L=D(G; E)$ for some system $E$ of endomorphisms of $G$, then $E\leqq K(G; L)$ and $L=D(G; K(G; L))$, though in general it may happen that $L<D(G; K)$.

The system $L$ of subgroups of $G$ shall be termed a *ring of subgroups*, if $L$ contains 0, $G$ and the cross-cuts and the sums (=join-groups) of each of its subsets. It is well known that rings of subgroups are Dedekind sets; and their importance for us lies in the fact that the sets $D(G; E)$ are rings of subgroups.

If $L$ is a ring of subgroups of the commutative group $G$, and if $X$ is a subset of $G$, then the cross-cut of all the subgroups in $L$ which contain $X$ is a subgroup in $L$; and we call this subgroup the $L$-subgroup $XL$ of $G$ or the *$L$-subgroup generated by $X$*. If $Z$ is an *$L$-subgroup* of $G$ such that the $L$-subgroups contained in $Z$ form a cycle (in the meaning of §I.2), then $Z$ is called a *cycle* in $L$; if the $L$-subgroup $Z$ is a cycle different from 0 in $L$, then $Z$ contains elements which are not contained in any proper $L$-subgroup of $Z$. If $z$ is such an element in $Z$, then $Z=zL$ so that we may say that cycles are cyclic. Since in general not every cyclic subgroup is a cycle, we define: The ring $L$ of subgroups of $G$ is *primary*, if every cyclic subgroup $gL$ in $L$ is a cycle in $L$.

If $S$ and $T<S$ are $L$-subgroups of $G$, then the $L$-subgroups $X$ between $T$ and $S$ define the $L$-subgroups $X/T$ of $S/T$. If $L$ is a primary ring of subgroups of $G$, then the $L$-subgroups of $S/T$ form a primary ring of subgroups too.

If $L$ is a primary ring of subgroups of $G$, $Z$ a cycle in $L$, then either $Z=0$ or $Z$ contains a uniquely determined subcycle $Z^*$ of order 1 (in $L$), a notation which we shall use occasionally.

If $g$ is an element in $G$, $L$ a primary ring of subgroups of $G$, then $gL$ is a cycle; and thus we may define the *$L$-order of* $g$ by the equation $n(g)=n(gL)$, the order of the cycle $gL$ in $L$. It is an immediate consequence of Theorem I.2.1 that *the order of the element $g$ in $G$ does not exceed $n$, if $g$ is the sum of elements in $G$ whose orders do not exceed $n$.*

If $x$ and $y$ are two elements of $L$-order 1, then either $xL=yL$ or else the cross-cut of $xL$ and $yL$ is 0. In the latter case $x+y$ is not contained in $xL$ and not in $yL$ so that $xL+yL$ contains at least three subgroups of order 1. From this remark one infers readily that $G$ is primary in the Dedekind set $L$ (using the definition of §I.3), if $L$ is a primary ring of subgroups of $G$ ([14]).

We note finally that the elements $x, y, \cdots$ are termed ($L$-) *independent*, if the subgroups $xL, yL, \cdots$ are independent elements of the Dedekind set $L$ (that is, if the cross-cut of $xL$ and of the sum $yL+\cdots$ is 0 and so on).

The following general theorem contains as special cases both the existence of the coordinates and the existence of the semi-linear transformations of a projective geometry.

([14]) Whether or not a ring of subgroups of an abelian group, satisfying the conditions of Theorem I.3.7 is a primary ring as defined in this section, seems to be an open problem.

THEOREM II.1.1. *If $L$ is a primary ring of subgroups of the abelian group $G$, if $J$ is a primary ring of subgroups of the abelian group $H$, if $L$ either does not contain any cycle of order $n$ or at least three independent ones, if $p$ is a projectivity of $L$ upon $J$, if $g$ is an element in $G$ and $h$ an element in $(gL)^p$, then there exists a linear transformation $q$ of $G$ into $H$ such that*

(i) $g^q = h$,

(ii) $x^q$ *is for every $x$ in $G$ an element in* $(xL)^p$,

(iii) $n(x) - n(x^q) = n(g) - n(h)$ *for* $n(g) - n(h) \leqq n(x)$ *and* $x^q = 0$ *for* $n(x) < n(g) - n(h)$.

**Proof**[15]. We note first that $p$ maps cycles in $L$ upon cycles in $J$ and that $p$ preserves both order of cycles and independence of subgroups. In particular we have $0 = 0^p$ and $H = G^p$. If $X$ is any subset of $G$, then it will prove convenient to put $p(X) = (XL)^p$.

(II.1.1.1) *If $x$ and $y$ are two independent elements in $G$ such that $0 < n(x) \leqq n(y)$, then $x$ and $y - x$ are independent, $n(y-x) = n(y)$ and the order of the cross-cut of $yL$ and $(y-x)L$ is $n(y) - n(x)$.*

From the hypothesis it follows that $xL + yL$ is the direct sum of $xL$ and $yL$; and it is obvious that it equals $xL + (y-x)L = yL + (y-x)L$. It is a consequence of Theorem I.2.1 that the order of $y - x$ does not exceed $n(y)$. Since $yL$ and $(xL + yL)/(xL)$ are isomorphic, it follows that $(y-x)L$ is, modulo its cross-cut with $xL$, a cycle of order $n(y)$; and this shows that $n(y-x) = n(y)$ and that $y - x$ and $x$ are independent. Since $xL$ and $(xL + yL)/(yL)$ are isomorphic, $(y-x)L$ is, modulo its cross-cut with $yL$, a cycle of order $n(x)$; and the order of the cross-cut of $yL$ and $(y-x)L$ is therefore $n(y-x) - n(x) = n(y) - n(x)$.

(II.1.1.2) *If $x$ and $y$ are two independent elements in $G$ such that $0 < n(x) \leqq n(y)$, and if $t$ is an element in $p(y)$, then there exists one and only one element $f(x, t; y)$ in $p(x)$ such that $f(x, t; y) \equiv t$ modulo $p(x-y)$. (Note that this statement holds trivially true for $x = 0$ in which case $f(x, t; y) = 0$.)*

The cross-cut of $p(x)$ and $p(x-y)$ is 0, since the cross-cut of $xL$ and $(x-y)L$ is 0 by (II.1.1.1); and consequently there exists at most one solution of our congruence, since the difference of any two solutions would be an element in the cross-cut of $p(x)$ and $p(x-y)$.—Since $xL + yL = xL + (x-y)L$, it follows that $p(x) + p(y) = p(x) + p(x-y)$; and since $t$ is an element in $p(x) + p(y)$, and $J$ a ring of subgroups, it follows that $t = r + s$ for $r$ in $p(x)$, $s$ in $p(x-y)$ or $r \equiv t$ mod $p(x-y)$ so that $r = f(x, t; y)$ is the required solution of our congruence.

---

[15] The method used in this proof is an adaptation and extension of a method employed by us previously in proving a similar theorem; cf. R. Baer, American Journal of Mathematics, vol. 61 (1938), pp. 1–44, in particular Footnote 10.

(II.1.1.3) *If $x$ and $y$ are independent elements in $G$ such that $n(x) \leqq n(y)$, and if $t$ is an element in $p(y)$, then $f(x, t; y) = 0$ for $n(x) \leqq n(y) - n(t)$ and $n(f(x, t; y)) = n(x) - (n(y) - n(t))$ for $n(y) - n(t) \leqq n(x)$.*

The order of the cross-cut of $p(y)$ and $p(y-x)$ is $n(y)-n(x)$, since this is —by (II.1.1.1)—the order of the cross-cut of $yL$ and $(y-x)L$. Thus $t$ is an element in this cross-cut and therefore in $p(y-x)$, if $n(t) \leqq n(y)-n(x)$. But if $t$ is in $p(y-x)$, then $f(x, t; y) = 0$ by (II.1.1.2).—It follows from the definition of $f(x, t; y)$ that $tJ+p(x-y) = f(x, t; y)\, J+p(x-y) = d$. Hence $f(x, t; y) = 0$ implies that $t$ is contained in $p(x-y)$ and that therefore $n(t) \leqq n(y)-n(x)$. Thus $f(x, t; y) \neq 0$, if $n(y)-n(x) < n(t)$. Since $xL$ and $(x-y)L$ are independent, so are $p(x)$ and $p(x-y)$; and since $f(x, t; y)$ is an element not 0 in $p(x)$, $f(x, t; y)$ and $p(x-y)$ are independent, so that $n(f(x, t; y)) = n(d/p(x-y))$. But $d/p(x-y)$ is isomorphic to $tJ$ modulo its cross-cut with $p(x-y)$; and this cross-cut equals the cross-cut of $p(y)$ and $p(x-y)$, since $t$ is in $p(y)$, but not in $p(x-y)$. Thus it follows finally from (II.1.1.1) that $n(f, t; y)) = n(t) - (n(y) - n(x))$, as was to be shown.

(II.1.1.4) *If $x, y, z$ are independent elements in $G$ such that $n(x) \leqq n(y) \leqq n(z)$, then $(z-x)L$ is the cross-cut of $zL+xL$ and $(z-y)L+(y-x)L$.*

Clearly $(z-x)L$ is contained in the cross-cut $C$ of $zL+xL$ and $(z-y)L+(y-x)L$.—It is a consequence of (II.1.1.1) that $zL+xL$ is the direct sum of $xL$ and $(z-x)L$; and $(z-x)L < C$ is therefore equivalent to dependence of $xL$ and $C$. But this is impossible, since the cross-cut of $xL$ and $(y-x)L + (z-y)L$ is 0 as a consequence of (II.1.1.1) and our hypotheses; and thus $C = (x-z)L$.

(II.1.1.5) *If $x, y, z$ are three independent elements in $G$ such that $n(x) \leqq n(y) \leqq n(z)$, and if $t$ is an element in $p(z)$, then*

$$f(x, t; z) = f(x, f(y, t; z); y).$$

Since $f(x, f(y, t; z); y) - t = f(x, f(y, t; z); y) - f(y, t; z) + f(y, t; z) - t$ is an element in the cross-cut of $p(xL+zL)$ and $p((x-y)L+(y-z)L)$, and since this cross-cut is $p(x-z)$ by (II.1.1.4), it follows that the element $f(x, f(y, t; z); y)$ is contained in $p(x)$ and satisfies $f(x, f(y, t; z); y) \equiv t \bmod p(x-z)$. But it follows from (II.1.1.2) that $f(x, t; z)$ is the only element meeting these requirements.

(II.1.1.6) *If $x, y, z$ are elements in $G$ such that $xL+yL$ and $zL$ are independent and such that $n(x) \leqq n(y) \leqq n(z)$, and if $t$ is in $p(z)$, then*

$$f(x + y, t; z) = f(x, t; z) + f(y, t; z).$$

The proof of this statement will be effected in three steps.

A. $x$ and $y$ are independent.

In this case the three elements $x$, $y$, $z$ are independent. Then it follows from (II.1.1.1) that $y$ and $z-y$ are independent, that $n(z)=n(z-y)$ so that $(z-y)L$ and $xL+yL$ are independent too. Likewise $y-x$ and $x$ are independent, $n(y)=n(y-x)$ so that the three elements $x$, $y-x$, $z-y$ are independent and satisfy $n(x)\leqq n(y-x)\leqq n(z-y)$. Hence it follows from (II.1.1.4) that the cross-cut of $xL+(z-y)L$ and $(z-x)L+yL=(x-y-x)L+(z-y-(x-y))L$ is just $(z-y-x)L$. Thus the element $v-t=(f(x, t; z)+f(y, t; z))-t$ is contained in $p(x+y-z)$, since it is contained in the cross-cut of $p(xL+(y-z)L)$ and $p(yL+(x-z)L)$. But $v=(v-t)+t$ is contained in the cross-cut $p(x+y)$ of $p(xL+yL)$ and $p((x+y-z)L+zL)=p((x+y)L+zL)$. Thus $v$ has been shown to be an element in $p(x+y)$ such that $v\equiv t$ mod $p(x+y-z)$; and this proves the required identity, since $f(x+y, t; z)$ is by (II.1.1.2) the only element satisfying these conditions.

B. $x=-y$.

We may assume $x\neq 0$, since $f(0, t; z)=0$. Then $xL+zL$ is the direct sum of $xL=yL$ and $zL$. Since there exist at least three independent elements of order $n(z)$ in $G$, we may infer from Corollary I.3.4 that there exists an element $v$ of order $n(z)$ in $G$ such that $x=-y$, $v$, $z$ are three independent elements. It is a consequence of (II.1.1.1) that $y$ and $v+x$ are independent elements too, $n(v)=n(v+x)$. Hence it follows from A that

$$\begin{aligned} f(v, t; z) &= f(v+x+y, t; z) = f(v+x, t; z) + f(y, t; z) \\ &= f(v, t; z) + f(x, t; z) + f(y, t; z) \end{aligned}$$

or

$$0 = f(x, t; z) + f(y, t; z).$$

C. $x$ and $y$ are not independent.

We may assume that neither $x$ nor $y$ is 0, since otherwise nothing need be proved. Thus $xL$ and $yL$ are dependent cycles different from 0 so that they have the same subcycle of order $1: c^*=(xL)^*=(yL)^*$. Since both $z$, $x$ and $z$, $y$ are pairs of independent elements, it follows that $(zL)^*+c^*$ contains every subcycle of order 1 of $xL+zL$ and of $yL+zL$.

If the three elements $x+y$, $y$, $z$ are independent, then $x+y$, $-y$, $z$ are independent too; and it follows from A and B that

$$\begin{aligned} f(x, t; z) &= f(x+y-y, t; z) = f(x+y, t; z) + f(-y, t; z) \\ &= f(x+y, t; z) - f(y, t; z). \end{aligned}$$

Thus we need only handle the case where $x+y\neq 0$ and $x+y$ and $y$ are dependent so that as before $((x+y)L)^*=c^*$. As under B there exists an element $v$ in $G$ such that $n(v)=n(z)$ and such that $v$ and $c^*+zL$ are independent. Then the triplets of elements $x$, $v$, $z$ and $x+y$, $v$, $z$ are triplets of independent elements whose orders do not exceed $n(z)$. Since $c^*$ is the only subcycle of order 1, contained in both $xL+vL$ and $yL+zL$, and since $x$ and $v+x$ are independent, it follows that $y$, $v+x$, $z$ is another triplet of independent elements

whose orders do not exceed $n(z)$. Thus it follows from A that $f(v, t; z)+f(x+y, t; z)=f(x+y+v, t; z)=f(x+v, t; z)+f(y, t; z)=f(v, t; z)+f(x, t; z)+f(y, t; z)$, completing the proof of the identity (II.1.1.6).

(II.1.1.7) *If $x$ and $z$ are two independent elements such that $n(x)\leqq n(z)$, and if $s$ is an element in $p(x)$, then there exists an element $t$ in $p(z)$ such that $s=f(x, t; z)$.*

Since $p(xL+zL)=p((x-z)L+zL)=p(x-z)+p(z)$, and since $s$ is an element in $p(xL+zL)$, there exist elements $r$, $t$ in $p(x-z)$ and $p(z)$, respectively, such that $s=r+t$. Then $s$ is an element in $p(x)$ such that $s\equiv t \bmod p(x-z)$; and since $t$ is in $p(z)$, it follows from (II.1.1.2) that $s=f(x, t; z)$.

We denote by $G_n$ the set of all those elements in $G$ whose $L$-order does not exceed $n$. It has been pointed out at the beginning of this section that $G_n$ is a subgroup in $L$. The set $H_n$ of all the elements in $H$ whose $J$-order does not exceed $n$ is likewise a $J$-subgroup of $H$.

(II.1.1.8) *If $z$ is an element of order $n$ in $G$, and if $t$ is in $p(z)$, then there exists one and only one linear transformation $f$ of $G_n$ into $H_n$ such that $z^f=t$, $x^f$ is in $p(x)$ for every $x$ in $G_n$, $n(x^f)-n(x)=n(t)-n$ for $n-n(t)\leqq n(x)$, but $x^f=0$ for $n(x)\leqq n-n(t)$.*

There exist by the hypotheses of Theorem II.1.1 two elements $x$, $y$ of order $n$ such that $x$, $y$, $z$ are three independent elements. We put $r=f(x, t; z)$ and $s=f(y, t; z)$. Since $t$ is an element in $p(z)$ such that $t\equiv r \bmod p(z-x)$, it follows from $n(x)=n$ and (II.1.1.2) that $t=f(z, r; x)$ and likewise that $t=f(z, s; y)$. Applying (II.1.1.5) we find that $f(x, t; z)=f(x, f(y, t; z); y)=f(x, s; y)$ and likewise $f(y, t; z)=f(y, r; x)$. It is a consequence of (II.1.1.3) that $n(r)=n(s)=n(t)$.

Any element $v\neq 0$ in $G_n$ belongs to one and only one of the following three classes.

*Class* 1. $v$ is independent of each of the three subgroups $xL+yL$, $yL+zL$ and $zL+xL$.

*Class* 2. $v$ is independent of two of the three subgroups $xL+yL$, $yL+zL$, $zL+xL$ and depends on the third one; and $v$ is independent of each of the three subgroups $xL$, $yL$, $zL$.

*Class* 3. $v$ is independent of one and only one of the three subgroups $xL+yL$, $yL+zL$, $zL+xL$; and $v$ is dependent of one and only one of the subgroups $xL$, $yL$, $zL$. (Thus dependence of $v$ and $z$ would imply independence of $vL$ and $xL+yL$.) We have to prove that these three classes exhaust all the possibilities. If $v$ is independent of $xL+yL$, then $v$ is independent of both $x$ and $y$. If $v$ is independent of two of the three subgroups $xL+yL$, $yL+zL$ and $zL+xL$, then $v$ is therefore independent of each of the elements $x$, $y$, $z$.— If $v$ depends on both $xL+yL$ and $yL+zL$, then $(vL)^*$ is contained in the cross-cut $yL$ of these two subgroups so that $v$ and $y$ are dependent. But then $v$ is certainly independent of $xL+zL$.

It follows from (II.1.1.2) and this trichotomy that of the three functions $f(u, r; x)$, $f(u, s; y)$ and $f(u, t; z)$ at least two are defined for $u=v$; and it follows from (II.1.1.5) and the properties of $r, s, t$ that those of these functions which are defined for $u=v$ have the same value $v^f$ for $u=v$. If we put $0^f=0$, then the function $f$ is defined for all the elements in $G_n$; and it follows from (II.1.1.2) and (II.1.1.3) that $v^f$ is a uniquely determined element in $p(v)$ for every $v$ in $G_n$; and that $v^f=0$ for $n(v)\leqq n-n(t)$, $n(v^f)-n(v)=n(t)-n$ for $n-n(t)\leqq n(v)$.

If $v$ and $w$ are any two elements (not both 0) in $G_n$, then at least one of the elements $x, y, z$ is by Corollary I.3.4 independent of the subgroup $vL+wL$. If, for example, $x$ and $vL+wL$ are independent, then $x$ and each of the three elements $v, v+w, w$ are independent so that $f(v, r; x)$, $f(v+w, r; x)$ and $f(w, r; x)$ are well determined elements; and since the orders of these elements do not exceed the order $n$ of $x$, it follows from (II.1.1.6) that $(v+w)^f=f(v+w, r; x)$ $=f(v, r; x)+f(w, r; x)=v^f+w^f$; and thus $f$ is a linear transformation meeting all the requirements.

If $g$ is any linear transformation which meets the requirements of (II.1.1.8), and if $v$ and $w$ are two independent elements in $G_n$, $n=n(w)$, then $v^g$ is an element in $p(v)$ such that $v^g\equiv w^g \bmod p(v-w)$, since $v^g-w^g$ $=(v-w)^g$ is an element in $p(v-w)$. If $x, y, z$ are the three elements used in the construction of $f$, then $x^g=f(x, z^g; z)=f(x, t; z)=r$, $y^g=s$, $z^g=t$—by (II.1.1.2). If $v$ is any element in $G_n$, then $v$ is 0 or independent of at least one of the elements $x, y, z$; and if $v$ and $x$ are independent, then it follows from (II.1.1.2) that $v^g=f(v, r; x)=v^f$ so that $f=g$.

During the remainder of the proof it will be convenient to term a linear transformation $f$ of $G_n$ into $H_n$ *permissible*, if $x^f$ is for every element in $G_n$ an element in $p(x)$, if $S^f$ is a $J$-subgroup of $H_n$ for every $L$-subgroup $S$ of $G_n$, and if there exists an integer $m\geqq 0$ such that $x^f=0$ for $n(x)\leqq m$ and $n(x)-n(x^f)=m$ for $m\leqq n(x)$.

(II.1.1.9) *Every permissible linear transformation of $G_n$ into $H_n$ is induced by a permissible linear transformation of $G_{n+1}$ into $H_{n+1}$.*

If—as we may assume without loss in generality—$G_n<G_{n+1}$, then there exist at least three independent elements of order $n+1$ in $G$. Let $x, y$ be any pair of independent elements of order $n+1$ in $G$, let $z$ be an element of order $n$ in $xL$ and let $f$ be a permissible linear transformation of $G_n$ into $H_n$. Since $z$ and $y$ are independent, (II.1.1.7) implies the existence of an element $t$ in $p(y)$ such that $z^f=f(z, t; y)$. There exists by (II.1.1.8) one and only one permissible linear transformation $g$ of $G_{n+1}$ into $H_{n+1}$ such that $y^g=t$ (put $m=n+1-n(t)$); and as shown in the last paragraph of the proof of (II.1.1.8) we have $z^g=f(z, t; y)=z^f$. Since $g$ induces a permissible linear transformation of $G_n$ into $H_n$, it follows from (II.1.1.8) that $g$ and $f$ coincide on $G_n$, as was to be shown.

Our theorem is now an immediate consequence of (II.1.1.8) and (II.1.1.9), provided the orders of the elements in $G$ are bounded, that is, $G = G_j$ for some integer $j$.—If the orders of the elements in $G$ are not bounded, then our theorem is again an immediate consequence of (II.1.1.8) and (II.1.1.9), if one remembers that every element in $G$ is contained in almost every $G_n$.

THEOREM II.1.2. *If $L$ is a primary ring of subgroups of the abelian group $G$ such that $L$ contains either no subcycle of order $n$ or at least three independent ones, then $L$ is the set $D(G; E)$ of all the $E$-admissible subgroups of $G$ where $E$ is the ring $K(G; L)$ of all the $L$-admissible endomorphisms of $G$.*

**Proof.** If $x$ is any element in $yL$ for $y$ an element in $G$, then there exists by Theorem II.1.1 an endomorphism $f$ of $G$ such that $y^f = x$ and such that $g^f$ is in $gL$ for every $g$ in $G$. Clearly $f$ belongs to $E$; and thus we have shown that $dL = dE$ for every $d$ in $G$, a fact that immediately implies $L = D(G; E)$.

THEOREM II.1.3. *If $L$ is a primary ring of subgroups of the abelian group $G$ such that $L$ contains either no cycles of order $n$ or at least three independent ones*[16], *and if $J$ is a primary ring of subgroups of the abelian group $H$, then every projectivity of $L$ upon $J$ is induced by an isomorphism of $G$ upon the whole group $H$.*

**Proof.** If $p$ is a projectivity of $L$ upon $J$, and if $g$ is an element not 0 in $G$, then $gL$ and $(gL)^p$ are cycles of equal order and there exists therefore an element $h$ such that $(gL)^p = hJ$. Since $n(g) = n(h)$, there exists by Theorem II.1.1 a linear transformation $q$ of $G$ into $H$ with the following properties: $g^q = h$, $x^q$ is for every element $x$ in $G$ an element in $(xL)^p$ such that $n(x) = n(x^q)$. Since therefore $x^q = 0$ implies $n(x) = 0$, that is, $x = 0$, we see that $q$ is an isomorphism such that $S^q \leqq S^p$ for every $L$-subgroup $S$ of $G$. If $u$ is any element not 0 in $S^p$, then $uJ$ is a cycle in $J$ and there exists one and only one subcycle $Z$ of $S$ such that $Z^p = uJ$ (and clearly $n(Z) = n(u)$). There exists in $G$ an element $y$ such that $y$ and $Z$ are independent and such that $n(y) = n(Z) = n(u) = n$. Since $yL$ and $Z$ are independent cycles of order $n$, so are $y^qJ$ and $uJ$; and $y^q - u$ and $u$ are independent elements of order $n$ too. There exists a cycle $T$ of order $n$ in $L$ such that $T^p = (y^q - u)J$ and one verifies that $yL + Z$ is the direct sum of $T$ and $Z$. Hence there exist uniquely determined elements $t$ and $z$ in $T$ and $Z$, respectively, such that $y = t + z$. Since $t^q$ is an element in $(y^q - u)J$, $z^q$ an element in $uJ$, $z^q - u = y^q - u - t^q$ is an element in the cross-cut of $uJ$ and $(y^q - u)J$. But this cross-cut is 0, since the cross-cut of $T$ and $Z$ is 0; and hence we have shown that $u = z^q$, $S^q = S^p$; and this completes the proof of the theorem.

[16] That this condition is indispensable for the validity of the theorem may be seen from simple examples like the groups of order a prime number (though here the theorem would hold true at least for auto-projectivities), or the direct sum of two abelian groups of order $p$ not 2 or 3 (where the theorem would not hold true for auto-projectivities); cf. R. Baer, loc. cit., p. 31.

2. **The ideals of the ring of endomorphisms.** If $L$ is a ring of subgroups of the abelian group $G$, then the endomorphism $e$ of $G$ is said to be $L$-admissible, if $Se \leqq S$ for every subgroup $S$ in $L$; and the set $E = K(G; L)$ of all the $L$-admissible endomorphisms of $G$ is a ring. If on the other hand a ring $E$ of endomorphisms of the abelian group $G$ has been given, then a subgroup $S$ of $G$ is termed *$E$-admissible*, if $Se \leqq S$ for every $e$ in $E$; and the set $L = D(G; E)$ of all the $E$-admissible subgroups of $G$ is a ring of subgroups. If this ring $D(G; E)$ of subgroups of $G$ is a primary ring of subgroups (as defined in §II.1), then we say for short that $G$ is *primary over* $E$; and it is the object of this section to characterize the rings $E$ with primary $D(G; E)$ by inner properties[17]; and to analyze the relations between the ideals in $E$ and the subgroups (in particular: the cycles) in $D(G; E)$. With this in mind we define: the *ring $E$ is primary if*

(i) *$E$ contains a universal unit* 1;

(ii) *every right-ideal in $E$ is two-sided*;

(iii) *the two-sided ideals not* 0 *in $E$ form a (finite or infinite) descending chain (of the order type of part of the negative integers).*

Such a primary ring $E$ contains one and only one greatest two-sided ideal different from $E$ which we shall denote by $P = P(E)$. We derive first some simple properties of $E$ and $P$.

(iv) *An element in $E$ possesses an inverse in $E$ if, and only if, it is not contained in $P$.*

**Proof.** If the element $z$ in $E$ is not in $P$, then the right-ideal $zE$ is not part of $P$ so that $zE = E$ by (ii). There exists therefore an element $z'$ in $E$ such that $zz' = 1$. Since $z'$ cannot be in $P$, there exists likewise an element $z''$ such that $z'z'' = 1$; and thus we have $z = zz'z'' = z''$ or $zz' = z'z = 1$.—That elements in $P$ do not possess inverses, is obvious.

(v) *Every two-sided ideal not* 0 *in $E$ is a power of $P$ and a principal right-ideal; and* 0 *is the cross-cut of all the $P^i$.*

**Proof.** If $Q$ is a two-sided ideal different from 0 in $E$, then there exists one and only one greatest two-sided ideal $Q'$ which is a proper part of $Q$. There exists an element in $Q$ which is not contained in $Q'$; and if $q$ is any such

[17] G. Köthe, Mathematische Zeitschrift, vol. 39 (1934), pp. 31–44; T. Nakayama, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 719–723; K. Asano, Japanese Journal of Mathematics, vol. 15 (1939), pp. 231–253; vol. 16 (1939), pp. 1–36 have treated the following problem: Given an abstract ring $R$, to find the necessary and sufficient conditions such that every abelian group admitting the elements in $R$ as operators and satisfying the maximum and minimum condition for admissible subgroups is the direct sum of cyclic admissible subgroups and such that every admissible subgroup of a cyclic group over $R$ is itself cyclic Then it is possible to derive necessary conditions in a trivial fashion by the remark that $R$ is an abelian group over $R$. Clearly our problem is quite different, as we consider a pair: abelian group $G$, ring $E$ of endomorphisms of $G$; and ask for conditions on $G$, $E$ assuring a satisfactory theory for the group $G$ over $E$. Moreover we have to exclude completely those pairs $G$, $E$ where $G$ is cyclic over $E$ whereas we may omit the maximum and minimum condition for admissible subgroups.

element, then $qE$ is a right-ideal contained in $Q$, but not in $Q'$ so that $Q=qE$. Thus $QP=qEP=qP$. If $p$ is an element in $P$, then $qp$ is an element in $Q$. If $qp$ would not be an element in $Q'$, then $Q=qpE$ so that there would exist an element $r$ in $E$, satisfying $q=qpr$. Since $p$ is in $P$, $1-pr$ is not in $P$ and possesses therefore by (iv) an inverse $t$ so that $q=q(1-pr)t=0$, an impossibility which proves that $QP\leqq Q'$. If $x$ is an element in $Q'$, then $x$ is in $Q=qE$ so that there exists an element $y$ satisfying $x=qy$. If $y$ would not be in $P$, then there would exist—by (iv)—the inverse $z$ of $y$ so that $q=qyz=xz$ would be an element in $Q'$, an impossibility which proves that $Q'\leqq QP$. Thus we have shown that $QP=Q'$ of which equality (v) is an immediate consequence.

From (iv) one infers that $P=0$ if, and only if, $E$ is a (not necessarily commutative) field; and in general $E/P$ is a field. Thus it follows from (v) that either $P=0$ or $P$ is the one and only one prime ideal in $E$. Any element $p$ such that $P=pE$ shall be called a *prime* in $E$; and if $p$ is a prime in $E$, then every two-sided ideal different from 0 in $E$ has the form $P^i=p^iE$.

If the number of two-sided ideals in $E$ is finite, then there exists a smallest positive integer $m=m(E)$ such that $P^m=0$; and if the number of two-sided ideals in $E$ is infinite, then we put $m(E)=\infty$. In the first case we infer from (iv) that an element in $E$ is a zero-divisor if and only if, it is contained in $P$; and in the second case it may be shown that none of the elements in $E$ is a zero-divisor.

THEOREM II.2.1. *Suppose that the ring $E$ of endomorphisms of the abelian group $G$ is a primary ring.*

(1) *If $m(E)$ is finite, then the ring $D(G; E)$ of $E$-admissible subgroups of $G$ is a primary ring of subgroups and $m(E)$ is the maximum order of the cycles in $D(G; E)$.*

(2) *$D(G; E)$ is for infinite $m(E)$ a primary ring of subgroups of $G$ if, and only if, there exists to every element $g$ in $G$ an element $e\neq 0$ in $E$ such that $ge=0$; and if $D(G; E)$ is primary, then there exist cycles of every order in $D(G; E)$.*

(3) *The order of the cycle $xE$ in $D(G; E)$ is $i$ if, and only if, $P^i$ is the set of all the elements $e$ in $G$ such that $xe=0$.*

(4) *If $Z$ is a cycle of order $n$ in $D(G; E)$, and if $0\leqq i\leqq n$, then $ZP^i$ is the uniquely determined subcycle of order $n-i$ of $Z$ (in $D(G; E)$) and $P^i$ contains every element $e$ in $E$ such that $Ze\leqq ZP^i$.*

**Proof.** If $g$ is an element in $G$, $N(g)$ the set of all the elements $e$ in $E$ such that $ge=0$, then $N(g)$ is a right-ideal in $E$; and hence it follows from (ii) and (v) that either $N(g)=0$ or $N(g)=P^i$ for suitable $i<m(E)$. If $g\neq 0$, then $N(g)<E$ and therefore $N(g)\leqq P$. If $e$ is an element in $E$ such that $gE=(ge)E$ (for $g\neq 0$ in $G$), then there exists an element $e'$ in $E$ such that $g=gee'$ so that $1-ee'$ is in $N(g)$ and therefore in $P$; and this implies that neither $e$ nor $e'$ is in $P$. Applying (iv) it follows now immediately that

(*) $gE=(ge)E$ for $g\neq 0$ if, and only if, $e$ possesses an inverse in $E$.

If $S$ is an $E$-admissible subgroup of $gE$, then the set of all the elements $e$ in $E$ such that $ge$ is in $S$ is a right-ideal between $N(g)$ and $E$. If $J$ and $J'$ are right-ideals such that $N(g) \leqq J < J'$, then $gJ$ and $gJ'$ are $E$-admissible subgroups such that $gJ < gJ'$ as follows from (ii), (iv), (v) and (*). But now it is clear that $gE$ is a cycle of order $i$ in $D(G; E)$ if, and only if, $N(g) = P^i$. This proves (3). To derive (4) we need note now only that $PN(gP) = N(g)$.

From the fact that $gE$ is a cycle of order $i$ if, and only if, $N(g) = P^i$, we infer that $D(G; E)$ is primary if, and only if, either $m(E)$ is finite, or $m(E)$ is infinite, but there exists to every $g$ in $G$ an $e \neq 0$ in $E$ such that $ge = 0$.—If the maximum order of the cycles in the primary ring $D(G; E)$ of subgroups of $G$ is $k$, and if $e$ is an element in $P^k$, then it follows from (3) that $ge = 0$ for every $g$ in $G$; and this implies $e = 0$, since $e$ is an endomorphism of $G$. This completes the proof of (1) and (2).

THEOREM II.2.2. *If the ring $E$ of endomorphisms of the abelian group $G$ contains the identity element* 1, *if the ring $D(G; E)$ of $E$-admissible subgroups of $G$ is primary and contains at least two independent cycles of order $m$, but no cycles of higher order, then the ring $E$ is a primary ring and contains every $D(G; E)$-admissible endomorphism of $G$.*

**Proof.** If the $E$-admissible subgroup $uE$ of $G$ is a cycle of the maximum order $m$, and if the $E$-admissible subgroup $vE$ of $G$ is independent of $uE$, then $vE$ is a cycle of an order not exceeding $m$ and the smallest $E$-admissible subgroup $W$ of $G$ which contains $u$ and $v$ is the direct sum $W = uE + vE$ of these two cycles. The $E$-admissible subgroup $(u+v)E$ is a cycle of an order not exceeding $m$ in $D(G; E)$ and $W$ is the sum of the cycles $vE$ and $(u+v)E$. If $C$ is the cross-cut of $(u+v)E$ and $vE$, then $uE$, $W/vE$ and $(u+v)E/C$ are isomorphic cycles; and this implies that $C = 0$ and that $(u+v)E$ is a cycle of order $m$ which is independent of $vE$.—Suppose now that $e$ is an element in $E$ such that $ue = 0$. Then $(u+v)e = ve$ is an element in the cross-cut $C$ of $(u+v)E$ and $vE$; and thus it follows that $ue = 0$ implies $ve = 0$, if $uE$ is a cycle of order $m$, and if $uE$ and $vE$ are independent.

Assume now that $uE$ is a cycle of order $m$, that $e$ is an element in $E$ such that $ue = 0$ and that $g$ is some element in $E$. If $gE$ is independent of $uE$, then we have already shown that $ge = 0$. If the cross-cut of $uE$ and $gE$ is different from 0, then there exists—by our hypothesis—an element $s$ in $G$ such that $sE$ is a cycle of order $m$ which is independent of $uE$. Since $gE$ is a cycle, and since $uE$ and $gE$ have their uniquely determined subcycle of order 1 in common, it follows that $sE$ and $gE$ are independent. But we have shown already that $ue = 0$ implies $se = 0$ and that $se = 0$ implies $ge = 0$. Since $e$ is an endomorphism we have therefore proved:

If $uE$ is a cycle of order $m$, and if $e$ is an element in $E$ such that $ue = 0$, then $e = 0$.

If $0 \leqq i \leqq m$, then denote by $(uE)^i$ the uniquely determined subcycle of

order $m-i$ of the cycle $uE$ of order $m$ and by $P_i$ the set of all the elements $e$ in $E$ such that $ue$ is in $(uE)^i$. Clearly every $P_i$ is a right-ideal in $E=P_0$; and we have just proved that $P_m=0$. If $J$ is any right-ideal in $E$, then $uJ$ is an $E$-admissible subgroup of $G$ so that $uJ=(uE)^j=uP_j$ for some $j$; and since $ue=uf$ for $e$ and $f$ in $E$ implies $e=f$, it follows now that $J=P_j$ so that every right-ideal in $E$ is contained in the descending chain of the $m+1$ right-ideals $P_j$. If $e$ is any element in $E$, then $eP_i$ is a right-ideal in $E$ and is hence a $P_j$ too. But $P_i<P_j$ would imply $P_i<eP_i<eP_j=e^2P_i< \cdots$ contradicting the fact that there exists but a finite number of right-ideals between $P_i$ and $E$. Thus $eP_i \leqq P_i$ so that the $P_i$ are two-sided ideals and this shows the primarity of the ring $E$.

Denote now by $F$ the ring of all the $D(G; E)$-admissible endomorphisms of $G$. Then $D(G; E)=D(G; F)$; and thus it follows from what we have shown already that $E \leqq F$, that $F$ is a primary ring and that $uf=0$ implies $f=0$, if $uF=uE$ is a cycle of order $m$. If $w$ is any element in $F$, then $uw$ is in $uE$. Hence there exists an element $e$ in $E$ such that $uw=ue$; and this implies $e=w$, since $uE=uF$ is a cycle of order $m$. Thus $E=F$; and this completes the proof.

THEOREM II.2.3. *If the ring $E$ of endomorphisms of the abelian group $G$ contains every $D(G; E)$-admissible endomorphism of $G$, if $D(G; E)$ is a primary ring of subgroups of $G$ which contains at least two independent cycles of every order $n$, then $E$ is a primary ring with the following property*(18).

(vi) *If $P$ is the prime ideal of $E$, if $e_i$ is an element in $P^i$ for $i=0, 1, 2, \cdots$, then there exists one and only one element $e$ in $E$ such that $e-e_0-e_1- \cdots -e_i$ is an element in $P^{i+1}$ (for every $i$).*

**Proof.** We denote by $G(n)$ the smallest $E$-admissible subgroup of $G$ which contains all the elements $x$ in $G$ such that $xE$ is a cycle of an order not exceeding $n$ (in $D(G; E)$); and we denote by $P(n)$ the set of all the elements $e$ in $E$ such that $G(n)e=0$. Clearly $P(n)$ is a two-sided ideal in $E$. If $x$ is any element in $G(n)$, then $x=x_1+ \cdots +x_j$ where the $x_i$ are elements in $G$ such that the $E$-admissible subgroup $x_iE$ is a cycle of an order not exceeding $n$ in $D(G; E)$. Since $xE$ is a subcycle of the sum of the cycles $x_iE$, it follows from Theorem I.2.1 that the order of the cycle $xE$ does not exceed $n$ either; and thus we have shown that the order of the cycle $xE$ in $D(G; E)$ does not exceed $n$ if, and only if, $x$ is in $G(n)$.

Since every element $x$ in $G$ is contained in some $G(n)$, it follows that the cross-cut of the descending chain of two-sided ideals $P(n)$ is 0. If $e_i$ is for $i=0, 1, 2, \cdots$ an element in $P(i)$, and if $a_i=e_0+ \cdots +e_i$, then all the endomorphisms $a_n, a_{n+1}, \cdots$ induce the same endomorphism $b_n$ in $G(n)$. Since every element in $G$ is contained in some $G(n)$, there exists therefore one (and only one) endomorphism $e$ of $G$ which coincides with $b_n$ on $G(n)$ (for every $n$).

(18) Because of this property (vi) the ring $E$ may be termed a $P$-adic ring.

Since every $a_n$ is $D(G; E)$-admissible, so is $e$; and hence $e$ is an element in $E$ such that $e-a_i$ is in $P(i+1)$.

Suppose that $s$ is any endomorphism in $E$. If $s\neq 0$, then there exists a greatest $n$ such that $s$ is in $P(n)$ (so that $s$ is not in $P(n+1)$) since 0 is the cross-cut of the $P(i)$. Suppose that $t$ is some element in $P(n)$ (which may or may not be in $P(n+1)$). Clearly $E(h)=E/P(h)$ is a ring of endomorphisms of $G(h)$ such that $D(G(h); E(h))$ consists only of the $E$-admissible subgroups of $G(h)$. Thus it follows from Theorem II.2.2 that $E(h)$ is a primary ring of endomorphisms of $G(h)$ whose only right-ideals are by (v) the two-sided ideals $P(i)/P(h)=(P(1)/P(h))^i$ for $0\leqq i\leqq h$. Thus there exists an element $s_h$ in $E$ such that $t-ss_h$ is an element in $P(n+h)$ (for $h=1, 2, \cdots$). Hence $s(s_h-s_{h+1})$ is in $P(n+h)$; and since $s$ is in $P(n)$, but not in $P(n+1)$, it follows that $(P(n+h)+s)E(n+h)=P(n)/P(n+h)$ and that therefore $s_h-s_{h+1}=e_h$ is in $P(h)$. Hence it follows from what has been shown in the preceding paragraph of this proof that there exists one and only one endomorphism $e$ in $E$ such that $e-e_0-\cdots-e_i$ is in $P(i+1)$. Then $t-s(s_0-e)=t-ss_i+s(s_i-s_0-e)$ $=t-ss_i+s(e_0+\cdots+e_{i-1}-e)$ is the sum of two elements in $P(n+i)$ so that $t-s(s_0-e)$ is part of the cross-cut of the $P(n+i)$ and is therefore 0. Consequently $t=s(s_0-e)$; or $P(n)=sE$; and this shows that every right-ideal not 0 in $E$ is one of the two-sided ideals $P(n)$. Hence $E$ is a primary ring of endomorphisms of $G$; and that $E$ satisfies (vi) has been shown in the second paragraph of the proof.

Theorem II.2.4. *If the ring $E$ of endomorphisms of the abelian group $G$ contains every $D(G; E)$-admissible endomorphism, if $D(G; E)$ is a primary ring of subgroups of $G$ which contains either no subcycle of order $n$ or at least two independent ones, then the following condition is necessary and sufficient for every left-ideal in $E$ to be two-sided.*

(vii) *If $U$ and $V$ are $E$-admissible subgroups of $G$ such that $V$ is the sum of $U$ and of a finite number of cycles in $D(G; E)$, and if there exists at most one cycle of order* 1 *in $V/U$, then $V/U$ is a cycle.*

**Proof.** If $P=P(E)=0$, then $E$ is a field and every cycle not 0 is of order 1 so that we may assume in the course of the proof that $P\neq 0$.—Suppose first that every ideal in $E$ is two-sided, and the $E$-admissible subgroups $U$ and $V$ meet the requirements of (vii). Then (vii) will be proved as soon as we have shown that there do not exist different cycles of equal order in $V/U$. We note first that the cycle $(U+x)E$ in $V/U$ is of order $n$—on account of the preceding theorems—if, and only if, the cross-cut of $U$ and $xE$ is exactly $xP^n$. If $(U+x)E$ and $(U+y)E$ are cycles of order $n$ in $V/U$, then $(U+x)P^{n-1}$ and $(U+y)P^{n-1}$ are cycles of order 1 in $V/U$ and hence it follows from the requirements on $V$ and $U$ enunciated in (vii) that $(U+x)P^{n-1}=(U+y)P^{n-1}$. If $p$ is any prime in $E$, then this implies the existence of an element $r$ in $E$ such that $U+xp^{n-1}=U+yp^{n-1}r$; and $r$ cannot be an element in $P$ since $xp^{n-1}$ is not

in $U$, though $yp^{n-1}P$ is contained in $U$. Since every left-ideal is assumed to be a right-ideal, there exists an element $s$ in $E$ such that $sp^{n-1}=p^{n-1}r$; and $s$ cannot be in $P$, since $p^{n-1}r$ is not in $P^n$. Since therefore $(x-ys)p^{n-1}=xp^{n-1}-yp^{n-1}r$ is an element in $U$, $((x-ys)E+U)/U$ is a cycle of an order not exceeding $n-1$. If we make now the induction hypothesis that there exists at most one subcycle of order $n-1$ of $V/U$, then it follows that $U+(x-ys)E$ is part of $U+yE$ so that $U+xE\leqq U+yE$; and consequently there exists at most one subcycle of order $n$ of $V/U$. Thus (vii) is a consequence of the fact that all the ideals in $E$ are two-sided.

Suppose now that (vii) is satisfied by $D(G; E)$, that $r$ is an element in $E$, though not in $P$, and that $p$ is a prime in $E$ (so that every right-ideal different from 0 in $E$ is of the form $p^iE$). If there exist cycles of order $n$ in $D(G; E)$, then there exist two independent cycles $xE$ and $yE$ of order $n$ in $D(G; E)$. If $U=(xp-ypr)E$ and $V=xE+yE$, then $U\leqq xP+yP$ so that $V/U$ is not a cycle, since $V/(xP+yP)$ is the direct sum of two cycles of order 1. Since the cross-cut of $U$ and $xE$ is null, it follows that $(U+xP^{n-1})/U$ is a cycle of order 1. Hence it follows from (vii) that there exists a cycle $(U+vE)/U$ of order 1 in $V/U$ which is different from $(U+xP^{n-1})/U$. Thus $v$ is not in $U$, but $vp$ is in $U$ and the cross-cut of $U+xP^{n-1}$ and $U+vE$ is exactly $U$. Furthermore there exist elements $s$, $t$ in $E$ such that $v=xs+yt$. If $t$ would be in $P$, then $t=pf$ for $f$ in $E$ and consequently (since $r$ is not in $P$ and (iv) may be applied)

$$\begin{aligned} v &= xs + ypf = xs + ypr\,r^{-1}f \\ &= x(s + pr^{-1}f) - (xp - ypr)r^{-1}f \\ &\equiv x(s + pr^{-1}f) \bmod U; \end{aligned}$$

and this is impossible, since it would imply $U+vE\leqq U+xE$. Thus $t$ is not in $P$. Since $vp$ is in $U$, there exists an element $g$ in $E$ such that $xsp+ytp=vp=(xp-ypr)g=xpg-yprg$; and from the independence of $xE$ and $yE$ we may infer $xsp=xpg$ and $ytp=yprg$. Since $t$ is not in $P$, we have $yE=ytE$ and hence $(ytp)E=yP=(yprg)E$ or $P=prgE$; and thus $g$ cannot be in $P$. Hence $xP=xpgE=xspE$ or $P=spE$ so that $s$ cannot be in $P$ either. Thus it follows from $xsp=xpg$ that $sp-pg$ is in $P^n$, since $n$ is the order of $xE$; and hence it follows from (iv) that $pg^{-1}\equiv s^{-1}p \bmod P^n$; and from $ytp=yprg$ we obtain likewise that $pr\equiv tpg^{-1}\equiv ts^{-1}p \bmod P^n$. Thus we have obtained the following intermediary result.

(*) If $r$ is in $E$ though not in $P$, if $p$ is a prime in $E$, and if there exist cycles of order $n$ in $D(G; E)$, then there exists an element $q(n)$ in $E$ such that $pr\equiv q(n)p \bmod P^n$.

If the orders of the cycles in $D(G; E)$ are bounded, and if $m$ is the maximum order of the cycles in $D(G; E)$, then $P^m=0$ and (*) implies $pr=q(m)p$.— If the orders of the cycles in $D(G; E)$ are not bounded, then it follows from (*) that $q(n)p\equiv q(n+1)p \bmod P^n$ so that $q(n)\equiv q(n+1) \bmod P^{n-1}$ for every $n$;

and hence it follows from Theorem II.2.3, (vi) that there exists one and only one element $q$ in $E$ such that $q \equiv q(n) \bmod P^{n-1}$ or $pr \equiv q(n)p \equiv qp \bmod P^n$ for every $n$. Since the cross-cut of the ideals $P^n$ is 0, it follows now that $pr = qp$; and thus we have shown

(**) If $r$ is in $E$ though not in $P$, if $p$ is a prime in $E$, then there exists an element $q$ in $E$ such that $pr = qp$.

Since every right-ideal different from 0 is a power of $P$ and of the form $p^iE$, every element not 0 in $E$ has the form $p^ir$ for $p$ a prime in $E$ and $r$ not in $P$. If $p^js$ is another element in this normal form, then there exist elements $q, t$ in $E$—by (**) and $sp^iE = p^iE$—such that $p^js\, p^ir = p^j\, p^i\, trs^{-1}\, s = p^iq\, p^js$ so that every left-ideal in $E$ is a right-ideal; and this completes the proof of our theorem.

It is a consequence of this theorem and the other theorems of this section, that by Theorem I.3.6 finite sums of cycles in the ring $D(G; E)$ are completely splitting, primary elements in a Dedekind set, if $E$ is a primary ring all of whose ideals are two-sided, and if there exist either no cycles of order $n$ in $D(G; E)$ or at least two independent ones.

If the ideals in the primary ring $E$ are two-sided, then it is readily verified that the subsets $GP^i$ of the group $G$ are $E$-admissible subgroups (do not only generate $E$-admissible subgroups); and on the basis of this remark one may prove by the customary arguments:

A([19]). *If $E$ is a primary ring of endomorphisms of the abelian group $G$ such that $m(E)$ is finite and such that all the ideals in $E$ are two-sided, then $G$ is a direct sum of cycles in $D(G; E)$.*

B([20]). *If the primary ring $E$ of endomorphisms of the abelian group $G$ contains every $D(G; E)$-admissible endomorphism of $G$, if every ideal in $E$ is two-sided, and if $D(G; E)$ is primary, then every $E$-admissible subgroup $S$ different from 0 of $G$ satisfying $S = SP$ is a direct summand of $G$ and is the direct sum of subgroups*([21]) *in $D(G; E)$ which contain one and only one cycle of order $n$ for every $n$.*

In order to show the independence of condition (vii) from the other conditions we construct an *example of a primary ring $E$ not all of whose ideals are two-sided.*

Let $F$ be a commutative field which possesses an isomorphism $v$ upon a proper subfield $F^v < F$. Consider the set $E$ of all the (ordered) pairs $(f, g)$ for $f$

([19]) See, for example, R. Baer, Compositio Mathematica, vol. 1 (1934), pp. 274–275. Note that this Theorem A is a special case of Theorem I.5.1 above.

([20]) See R. Baer, Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 800–806.

([21]) H. Prüfer has introduced subgroups of this type into the study of primary abelian groups; in analogy to Prüfer's terminology they may be called "groups of type $P^\infty$" or "cycles of order $\infty$."

and $g$ in $F$. Two such pairs $(f, g)$ and $(f', g')$ define the same element in $E$ if, and only if, $f=f'$, $g=g'$; their sum is defined by $(f, g)+(f', g')=(f+f', g+g')$ and their product by $(f, g)(f', g')=(ff', f^vg'+gf')$. One verifies readily that $E$ is a ring with identity $1=(1, 0)$ and $0=(0, 0)$, that $(f, g)$ possesses an inverse in $E$ if, and only if, $f\neq 0$, and that therefore the only right-ideal different from 0 and $E$ is the set of all the elements $(0, f)$ for $f$ in $F$, since $(0, f)(g, 0)=(0, fg)$. This ideal $(0, F)$ is clearly a two-sided ideal whose square is 0 so that $E$ is a primary ring. A left-ideal is formed by all the elements $(0, f^v)$ for $f$ in $F$, as follows from the above formulas and the multiplicativity of $v$. Since $0<F^v<F$, this left-ideal is different from all the right-ideals([22]).

If in the preceding construction we would choose $F$ as a finite field possessing an automorphism $v\neq 1$ (so that $F=F^v$), then $E$ would be *a finite, primary, noncommutative ring all of whose ideals are two-sided.*

**Application to the principles of projective geometry.** If $L$ is a primary ring of subgroups of the abelian group $G$ such that the orders of the cycles in $L$ do not exceed 1 (so that cycles not 0 are *points*) and such that there exist at least three independent cycles of order 1 in $L$, then it is a consequence of Theorem II.1.2 that $L$ is the ring of all the $E$-admissible subgroups of $G$ where $E$ is the ring of all the $L$-admissible endomorphisms of $G$; and it is a consequence of Theorem II.2.2 that the ring $E$ is a primary ring of endomorphisms of $G$. Since the maximum order of the cycles in $L$ is 1, it follows from Theorem II.2.1 and (iv) that $E$ is a (not necessarily commutative) field. But this implies immediately that Desargues' theorem is valid in the projective geometry represented by $L$. Since the possibility of representing the linear subspaces of a projective geometry by means of a ring of subgroups of an abelian group may be considered to be the essence of representations by means of homogeneous coordinates, we may state this result somewhat loosely as follows.

*A projective plane admits of a representation by means of homogeneous coordinates if, and only if, Desargues' theorem holds true in it.*

3. **The fundamental theorem of projectivity.** If $E$ is a ring of endomorphisms of the abelian group $G$, and if $f$ is an isomorphism of $G$ upon the (whole) abelian group $H$, then there exists the inverse isomorphism $f^{-1}$ of $f$ which maps $H$ upon $G$; and mapping the endomorphism $e$ in $E$ upon $e^f=f^{-1}ef$ constitutes an isomorphism of $E$ upon a ring of endomorphisms of $H$ which we call the isomorphism of $E$ induced by $f$.

THEOREM II.3.1([23]). *If $E^i$ is a primary ring of endomorphisms of the abelian group $G^i$ $(i=1, 2)$, if the ring $D(G^i, E^i)$ of the $E^i$-admissible subgroups of $G^i$ is primary and contains either no cycles of order $n$ or at least three independent ones,*

([22]) There are many possibilities of generalizing this construction.

([23]) This theorem asserts—in the terminology of projective geometry—that every projectivity is induced by a semi-linear transformation.

*if $E^i$ contains every $D(G^i, E^i)$-admissible endomorphism of $G^i$, and if $p$ is a projectivity of $D(G^1, E^1)$ upon $D(G^2, E^2)$, then there exists an isomorphism $q$ of $G^1$ upon $G^2$ which induces $p$ in $D(G^1, E^1)$ and which induces an isomorphism of $E^1$ upon $E^2$.*

The existence of an isomorphism $q$ of $G^1$ upon $G^2$ which induces $p$ in $D(G^1, E^1)$ is an immediate consequence of Theorem II.1.3. Furthermore we know that such an isomorphism $q$ induces an isomorphism of $E^1$ upon a ring $F$ of endomorphisms of $G^2$; and for reasons of symmetry it suffices to show that $F \leqq E^2$. Thus if $e$ is an endomorphism of $G^1$, $S$ an $E^2$-admissible subgroup, then $Se^q = (S^{q^{-1}}e)^q = (S^{p^{-1}}e)^p \leqq (S^{p^{-1}})^p = S$ so that $e^q$ is $D(G^2, E^2)$-admissible and therefore in $E^2$.

If in particular $G = G^1 = G^2$, $E = E^1 = E^2$, then $p$ is a projectivity of $D(G; E)$ (upon itself), $q$ an automorphism of $G$ and $q$ induces an automorphism in $E$. If $g$ is an element in $G$, $e$ an endomorphism in $E$, then $(ge)^q = ((g^q)^{q^{-1}}e)^q = g^q e^q$; and it is easily verified that exactly those automorphisms of $G$ which induce an automorphism in $E$ induce a projectivity of $D(G; E)$ (upon itself). In analogy to the distinction between linear and quasi-linear transformations we say that the automorphism $q$ of $G$ is a *proper E-automorphism*([24]), if $q$ induces the identity in $E$, that is, if $q$ commutes with all the endomorphisms in $E$.

We consider now—throughout this section—an abelian group $G$, a primary ring $E$ of endomorphisms of $G$ such that every left-ideal (as well as right-ideal) in $E$ is two-sided and such that the primary ring $D(G; E)$ of subgroups of $G$ is the sum of a finite number of cycles in $D(G; E)$. It is then a consequence of Theorem II.2.1, Theorem II.2.4—using the additional hypothesis that $G$ contains at least two independent cycles of maximum order $m = m(E)$—and Theorem I.3.7 that $G$ splits completely and is primary in the Dedekind set $D(G; E)$. Consequently $G$ is the direct sum of a finite number of cycles $Z_1, \cdots, Z_k$ in $D(G; E)$; and it follows from Lemma I.3.8 that there exists a cycle $Z$ in $D(G; E)$ which is not part of any proper partial sum of the $Z_i$; it is readily verified that this is equivalent to saying that $G = Z + \sum_{j \neq i} Z_j$ for every $i$.

If $G$ is the direct sum of the cycles $Z_{1i}$ and of the cycles $Z_{2i}$ in $D(G; E)$, then it follows from Corollary I.3.4 that the numbering may be effected in such a way that $n(Z_{1i}) = n(Z_{2i})$ for every $i$. If furthermore $Z_i$ is a cycle such that $G = Z_i + \sum_{j \neq h} Z_{ih}$ for every $j$, then our generalization of the fundamental theorem of projectivity may be stated as follows([25]).

THEOREM II.3.2. *There exists one and only one projectivity $p$ of $D(G; E)$ which is induced by a proper E-automorphism of G and which satisfies: $Z_{1j}^p = Z_{2j}$, $Z_1^p = Z_2$.*

---

([24]) It is customary in the theory of abelian operator groups to admit only these "proper $E$-automorphisms" as automorphisms.

([25]) If one takes into account Theorem II.3.3. below.

**Proof.** To prove the unicity of $p$ we consider a proper $E$-automorphism $f$ of $G$ which leaves $Z_1$ and every $Z_{1j}$ invariant. Since the $Z$'s are cycles in $D(G; E)$, there exists an element $z$ in $Z_1$ such that $Z_1 = zE$; and since $G$ is the direct sum of the $Z_{1i}$, there exist uniquely determined elements $z_i$ in $Z_{1i}$ such that $z = z_1 + \cdots + z_k$. From the choice of $z$ and $Z_1$ it follows immediately that $Z_{1i} = z_iE$. From the choice of $f$ it follows $Z_1 = z^fE$, $Z_{1i} = z_i^fE$; and there exist therefore elements $e$, $e_i$ in $E$ (though not in the prime ideal $P$ of $E$) such that $ze = z^f$, $z_ie_i = z_i^f$. But now it follows immediately that $z_ie = z_ie_i$ so that $g^f = ge$ for every $g$ in $G$, that is, $f$ induces the identity in $D(G; E)$; and this proves the unicity of the required projectivity.

To prove the existence of some projectivity meeting the requirements we note first that there exist elements $z_{ij}$ such that $Z_{ij} = z_{ij}E$ and such that $Z_i = (z_{i1} + \cdots + z_{ik})E$. Since $n(z_{1j}) = n(z_{2j})$ for every $j$, there exists one and only one proper $E$-automorphism $q$ such that $z_{1j}^q = z_{2j}$; and $q$ clearly induces a projectivity $p$ of $D(G; E)$ (upon itself) such that $Z_{1j}^p = Z_{2j}$, $Z_1^p = Z_2$.

The proper $E$-automorphism $f$ of $G$ shall be termed a *perspectivity of* $G$, if there exists an $E$-admissible direct summand $F$ of $G$ such that $G/F$ is a cycle and such that every element in $F$ is left invariant by $f$. That this definition([26]) is not too narrow may be seen from the following fact.

*The projectivity $p$ of $D(G; E)$ (upon itself) is induced by a perspectivity of $G$, if there exists an $E$-admissible direct summand $T$ of $G$ such that $G/T$ is a cycle and such that every $E$-admissible subgroup of $T$ is left invariant by $p$, provided $G$ contains at least three independent elements of maximum order.*

**Proof.** It is a consequence of Theorem II.3.1 that $p$ is induced in $D(G; E)$ by some automorphism $q$ of $G$ which induces an automorphism of $E$. It is a consequence from our general hypotheses that $T$ possesses a basis $B$ and $B$ contains certainly two different elements of maximum order $m$ in $G$. If $t$ is an element of order $m$ in $B$, and if $b$ is an element not $t$ in $B$, then $t^q = te$ for $e$ in $E$ though not in $P$, $b^q = be'$, $(b+t)^q = (b+t)e''$ for $e'$, $e''$ in $E$. Consequently $te = te''$, $be' = be''$. Since $t$ is of order $m$, $P^m = 0$, we have $e = e''$, $b^q = be' = be'' = be$; and this implies $x^q = xe$ for every $x$ in $T$. Since $e$ is not in $P$, there exists an inverse $e^{-1}$ of $e$ in $E$. If we put $y^f = y^qe^{-1}$ for every $y$ in $G$, then $f$ and $q$ induce the same projectivity $p$ in $D(G; E)$, but $f$ is a perspectivity, since it leaves every element in $T$ invariant, and since $tv = t^{f^{-1}}v = (t^{f^{-1}}v)^f = tv^f$ and $P^m = 0$ imply $v = v^f$ for $v$ in $E$.

THEOREM II.3.3. *The group of proper $E$-automorphisms of $G$ is generated by the perspectivities of $G$.*

---

([26]) It is readily verified that this definition and the customary definition—postulating a center apart from the "axis" we postulated—coincide in the case of ordinary projective geometry, provided one is able, as we are, to use Desargues' theorem.

**Proof**(27). The proper $E$-automorphism $f$ of $G$ shall be termed *irreducible* (in the course of this proof), if there exists an $E$-admissible direct summand $T$ of $G$ with the following properties:

(i) Every element in $T$ is left invariant by $f$.

(ii) If $f$ is the product of the proper $E$-automorphisms $u$ and $v$, if the $E$-admissible direct summands $U$ and $V$ of $G$ both contain $T$, and if $u$ leaves the elements in $U$, $v$ the elements in $V$ invariant, then $U=T$ or $V=T$.

It is an obvious consequence of the maximum condition for $E$-admissible subgroups, that every proper $E$-automorphism of $G$ is the product of irreducible proper $E$-automorphisms of $G$; and thus all we have to prove is the following statement.

*A proper $E$-automorphism of $G$ is a perspectivity if, and only if, it is irreducible.*

The irreducibility of perspectivities is a consequence of the fact that $S=G$, if $S$ is an $E$-admissible direct summand of $G$ such that there exists an $E$-admissible direct summand $T$ of $G$ satisfying $T<S$ and $G/T$ is a cycle.

Thus let us assume now that the proper $E$-automorphism $f$ of $G$ is irreducible. Then there exists an $E$-admissible direct summand $T$ of $G$ such that $f$, $T$ satisfy the above conditions (i), (ii). Since $T$ is a direct summand of $G$, there exists an element $b$ and an $E$-admissible subgroup $U$ of $G$ such that $G$ is the direct sum of $T$, $U$ and $bE$ and such that the orders of the elements in $U$ do not exceed $n(b)$.—If the elements $b$ and $b^f$ were independent modulo $T$, then there would exist an $E$-admissible subgroup $V$ of $G$ such that $G$ would be the direct sum of $T$, $V$, $bE$ and $b^fE$. Since $b$ and $b^f$ are of equal order, there exists one and only one proper $E$-automorphism $v$ of $G$ which leaves every element in $T+V$ invariant and which interchanges $b$ and $b^f$. This is impossible, since $v$ leaves every element in the direct summand $T+V+(b+b^f)E$ invariant, and since $fv^{-1}$ leaves every element in the direct summand $T+V+bE$ invariant. Thus it follows that $b$ and $b^f$ are dependent modulo $T$; and this implies that $G$ is the direct sum of $T$, $U$ and $b^fE$ too. Consequently there exists one and only one proper $E$-automorphism $u$ of $G$ which leaves every element in $T+U$ invariant and which maps $b$ upon $b^f$. Since $f$ is irreducible, and since $fu^{-1}$ leaves every element in $T+bE$ invariant, it follows that $T=T+U$ or $U=0$; and this implies that $f$ is a perspectivity.

If $D(G; E)$ contains at least three independent cycles of maximum order, then it follows from Theorem II.3.1 that every projectivity of $D(G; E)$ upon itself is induced by a proper $E$-automorphism of $G$ if, and only if, every automorphism of $E$ is an inner automorphism. Thus we see that projective geometry over the field of real numbers has—as far as the behaviour of projectivities is concerned—more in common with abelian groups of order a power of a prime than with projective geometry over the field of complex

(27) It should be noted that this proof is slightly simpler and proves more than the customary proofs of the projective special case of this theorem.

numbers, since both the field of real numbers and the ring of integers modulo a power of a prime admit of the identity-automorphism only, whereas the field of complex numbers possesses an infinity of automorphisms.

4. **Duality and the theory of characters.** Throughout this section we make the following *assumptions*. $E$ is a primary ring of endomorphisms of the abelian group $G$; every ideal in $E$ is a two-sided ideal in $E$; $G$ is in the ring $D(G; E)$ of the $E$-admissible subgroups of $G$ the sum of a finite number of cycles; $D(G; E)$ contains[28] at least two independent cycles of maximum order $m$. We note the following consequences of these hypotheses (and Theorems II.2.1, II.2.4 and I.3.7): if $P$ is the prime ideal in $E$ (unless $E$ is a field and $P=0$), then $0=P^m<P^{m-1}$; if $S$ is an $E$-admissible subgroup of $G$, then $S$ is the direct sum of a finite number of cycles in $D(G; E)$ and $G/S$ is the direct sum of a finite number of cycles in $D(G/S; E)$.

*A character*[29] *of $G$ in $E$* is a single-valued function $f$ of the elements in $G$ with values in $E$ such that $f(ge+g'e')=f(g)e+f(g')e'$ for $g$, $g'$ in $G$ and $e$, $e'$ in $E$. If $f$ and $v$ are characters of $G$ in $E$, and if $e$ is an element in $E$, then $f(g)+v(g)$ and $ef(g)$ are characters of $G$ in $E$; and thus it follows that the set $Ch(G; E)$ of all the characters of $G$ in $E$ is an abelian group, admitting the elements in $E$ as left-operators. Characters and character group of $Ch(G; E)$ in $E$ are defined in a like manner, apart from certain obvious interchanges of right and left.

THEOREM II.4.1. (a) *$G$ is essentially the same as the group of characters of $Ch(G; E)$ in $E$.* (b) *$D(G; E)$ and $D(Ch(G; E); E)$ are duals of each other.*

**Proof.** If $g$ is an element in $G$, $f$ an element in $Ch(G; E)$, then we put $Q_g(f)=f(g)$. It is readily verified that $Q_g$ is for every $g$ in $G$ a character of $Ch(G; E)$ in $E$, that $Q_{x+y}=Q_x+Q_y$, $Q_{xe}=Q_xe$ for $x$, $y$ in $G$, $e$ in $E$. To prove that $Q_g=0$ implies $g=0$; suppose that $g\neq 0$ and that $B$ is a basis of $G$ over $E$. Then $g=\sum_{b \text{ in } B} be(b)$ for $e(b)$ in $E$; and $g\neq 0$ implies that at least one $be(b)\neq 0$. If $p$ is a prime in $E$, then there exists one and only one character $s=s_b$ of $G$ in $E$ which maps $b$ upon $p^{m-n(b)}$ and all the other elements in $B$ upon 0. Thus $s(g)=p^{m-n(b)}e(b)$; and this is not 0, since $e(b)$ would otherwise be an element in $P^{n(b)}$ so that $be(b)$ would be 0.—Thus we have proved that an isomorphism of $G$ upon a group of characters of $Ch(G; E)$ in $E$ is established by mapping the element $g$ in $G$ upon the character $Q_g$ of $Ch(G; E)$. To prove that this isomorphism exhausts the character group of $Ch(G; E)$ we consider again the characters $s_b$ of $Ch(G; E)$ (for $b$ in some basis $B$ of $G$ over $E$). From the fact that every right- and every left-ideal in $E$ is of the form $Ep^i=p^iE=P^i$ we infer readily that the $s_b$ for $b$ in $B$ form a basis of $Ch(G; E)$ over $E$. If $v$ is a

[28] This last hypothesis is only needed in order to be able to apply Theorem II.2.4. After this one application has been effected, this hypothesis may be dropped.

[29] For generalizations of the concept of character, see P. Lewis, *Characters of abelian groups*, American Journal of Mathematics, vol. 64 (1942), pp. 81–105.

character of $Ch(G; E)$ in $E$, then $p^{n(b)}s_b = 0$ implies that $v(s_b) = p^{m-n(b)}d(b)$ for $d(b)$ in $E$. If $g = \sum_{b \text{ in } B} bd(b)$, then $Q_g(s_b) = s_b(g) = p^{m-n(b)}d(b) = v(s_b)$ so that $Q_g = v$; and this completes the proof of (a).

If $S$ is an $E$-admissible subgroup of $G$, then we denote by $(f(S) = 0)$ the set of all the characters of $G$ in $E$ which map $S$ upon 0; and the analogous definition may be used for $E$-admissible subgroups $S$ of $Ch(G; E)$.—If $S$ is an $E$-admissible subgroup of $G$, then every character of $G/S$ in $E$ is induced by one and only one character in $(f(S) = 0)$; and since we showed in the first part of the proof that 0 is the only element in $G/S$ mapped upon 0 by all the characters of $G/S$ in $E$, it follows that[30] $(f((f(S) = 0)) = 0) = S$. Since—by (a)—$G$ is the character group of $Ch(G; E)$, the same formula holds true for $E$-admissible subgroups of $Ch(G; E)$. But now it is readily verified that mapping the $E$-admissible subgroup $S$ of $G$ upon the $E$-admissible subgroup $(f(S) = 0)$ of $Ch(G; E)$ constitutes a biunivoque and monotonically decreasing correspondence (that is, a duality) between $D(G; E)$ and $D(Ch(G; E); E)$.

THEOREM II.4.2. *If $D(G; E)$ contains at least three independent cycles of maximum order, then the existence of an anti-automorphism of $E$ is a necessary and sufficient condition for the existence of an auto-duality of $D(G; E)$.*

REMARK. We note that consequently not every primary abelian operator group is self-dual.—Since there exists an auto-duality of $D(G; E)$ whenever $G$ is the direct sum of two cycles of order 1 in $D(G; E)$, the assumption of the existence of at least three independent cycles of maximum order is indispensable[31].

**Proof.** Every element $e$ in $E$ induces an endomorphism $e'$ of $Ch(G; E)$; since $Ch(G, E)$ contains elements of order $m$ (if $n(b) = m$, then the character $s_b$ constructed in the proof of Theorem II.4.1 is of order $m$) $e' = 0$ implies $e = 0$; since $(e+d)' = e' + d'$, $(ed)' = d'e'$ for $d$, $e$ in $E$, it follows that mapping $e$ upon $e'$ constitutes an anti-isomorphism of $E$ upon a ring $E'$ of endomorphisms of $Ch(G; E)$. It is a consequence of Theorems II.4.1, (b), II.2.1 and II.2.2 that $E'$ contains every $D(Ch(G; E); E)$-admissible endomorphism of $Ch(G; E)$.

If there exists an anti-automorphism of $E$, then there exists an isomorphism $f$ of $E$ upon $E'$ and there exists one and only one isomorphism $v$ of $G$ upon $Ch(G; E)$ which satisfies $b^v = s_b$ for $b$ in a basis $B$ of $G$ and $s_b$ defined as in the proof of Theorem II.4.1 and which satisfies furthermore $(ge)^v = g^v e^f$ for $g$ in $G$ and $e$ in $E$. Thus there exists a projectivity of $D(G; E)$ upon its—by Theorem II.4.1, (b)—dual $D(Ch(G; E); E)$, proving the self-duality of $D(G; E)$.

If there exists an auto-duality of $D(G; E)$, then there exists by Theorem

---

[30] This identity is in the projective special case essentially the content of the theory of linear equations.

[31] That the existence of a duality cannot be expected, unless $G$ is the sum of a finite number of cycles in $D(G; E)$, has been pointed out before; cf. R. Baer, Duke Mathematical Journal, vol. 5 (1939), pp. 824–838.

II.4.1, (b) a projectivity of $D(G; E)$ upon $D(Ch(G; E); E)$; and hence it follows from Theorem II.3.1 that $E$ and $E'$ are isomorphic. Since $E$ and $E'$ have been shown to be anti-isomorphic in the first paragraph of the proof, this implies the existence of an anti-automorphism of $E$.

5. **The theorem of Pappus.** Throughout this section we shall assume that $E$ is a primary ring of endomorphisms of the abelian group $G$, that $P$ is the greatest two-sided ideal different from $E$ in $E$, that the ring $D(G; E)$ of all the $E$-admissible subgroups of $G$ is primary and contains either no subcycles of order $n$ or at least three independent ones, and that $E$ contains every $D(G; E)$-admissible endomorphism of $G$.

The seven cycles $U_1, U_2, U_3 / Z / V_1, V_2, V_3$ in $D(G; E)$ are in *Pappus order*[32], if they are of equal order $n$, if $Z, U_1, V_1$ are independent, and if

$$U_1 + Z = U_2 + Z = U_3 + Z = U_1 + U_2 = U_2 + U_3,$$
$$V_1 + Z = V_2 + Z = V_3 + Z = V_2 + V_1 = V_1 + V_3.$$

Before stating the extension of Pappus' theorem whose proof is the goal of this section, we establish a useful normal form for cycles in Pappus order.

LEMMA II.5.1. *If the seven cycles $U_1, U_2, U_3 / Z / V_1, V_2, V_3$ of order $n$ in $D(G; E)$ are in Pappus order, if $W(i, j) = W(j, i)$ for $i \neq j$ is the cross-cut of $U_i + V_j$ and $U_j + V_i$, then the $W(i, j)$ are cycles of order $n$; and there exist three independent elements $z, u, v$ of order $n$ in $G$ and elements $1+x, y$ in $E$ though not in $P$ such that*

$$\text{(N)}\qquad \begin{aligned} &U_1 = uE, \quad U_2 = (u - z)E, \quad U_3 = (u(1 + x) - zx)E,\\ &Z = zE,\\ &V_1 = vE, \quad V_2 = (v - z)E, \quad V_3 = (vy + z)E,\\ &W(1, 2) = (u + v - z)E, W(2, 3) = ((v - z)y + (u(1 + x) - zx)(1 + y))E,\\ &W(3, 1) = (u(1 + x) - (vy + z)x)E. \end{aligned}$$

**Proof.** There exist elements $z$, $b$ in $G$ such that $Z = zE$, $U_1 = bE$. Then $U_2 = (zr + bs)E$ for $r$, $s$ in $E$. Since $U_2$ is part of $U_1 + Z$, but of no proper partial sum of this direct sum, neither $r$ nor $s$ can be in $P$—by Theorem II.2.1—so that $r^{-1}$ exists in $E$. Hence $U_1 = uE$ for $u = -bsr^{-1}$ and $U_2 = (z - u)E$. Likewise we find an element $v$ such that $V_1 = vE$, $V_2 = (z - v)E$. The independence of $z, u, v$ is a consequence of the independence of $Z, U_1, V_1$.—Since $U_3$ is part of the direct sum $Z + U_2$, but of no proper partial sum of $Z + U_2$, there exist ele-

(32) This arrangement of the cycles is necessitated by typesetting limitations. The more suggestive form is used in (N) below. Note the asymmetry in the treatment of 1 and 2, though it is possible to interchange 1 and 2, provided one interchanges at the same time $U$ and $V$. The customary form of stating the theorem of Pappus is so wide that there exists hardly a geometry in which it holds true and it will become apparent from the proof of Theorem II.5.2 below that the restrictions we imposed upon the cycles $Z$, $U_i$, $V_i$ are unavoidable.

ments $r$, $s$ in $E$, though not in $P$, such that $U_3=(zr+(z-u)s)E$. Let $x=-sr^{-1}-1$. Then $1+x$ is not in $P$ and $U_3=(u(1+x)-zx)E$.—Since $V_3$ is part of the direct sum $Z+V_1$, but of no proper partial sum of $Z+V_1$, there exist elements $d$, $t$ in $E$, though not in $P$ such that $V_3=(zd+vt)E$. Then $y=td^{-1}$ is not in $P$ and $V_3=(vy+z)E$.

The elements in $W(1, 2)$ are of the form $ur+(v-z)s=vh+(u-z)k$; since this equation implies $ur=uk$, $zk=zs$, $vs=vh$, it follows that $W(1, 2)=(u+v-z)E$.

The elements in $W(2, 3)$ are of the form $(u-z)r+(vy+z)s=(v-z)h+(u(1+x)-xz)k$; this equation implies that $r\equiv(1+x)k$, $ys\equiv h$, $r-s\equiv h+xk$ mod $P^n$, as follows from the independence of $z$, $u$, $v$ and of Theorem II.2.1; but these congruences imply $h\equiv ys$, $k\equiv r-xk\equiv h+s\equiv(y+1)s$ mod $P^n$ so that $W(2, 3)=((v-z)y+(u(1+x)-zx)(1+y))E$.

The elements in $W(3, 1)$ are of the form $ur+(vy+z)s=vh+(u(1+x)-zx)k$ and this implies $ur=u(1+x)k$, $vys=vh$, $zs=-zxk$ or $r\equiv(1+x)k$, $ys\equiv h$, $s\equiv -xk$ mod $P^n$ so that $W(3, 1)=(u(1+x)-(vy+z)x)E$.

Since $z$, $u$, $v$ are three independent elements of order $n$, it is now clear that the $W(i, j)$ are cycles of order $n$.

THEOREM II.5.2. *If $D(G; E)$ contains three independent cycles of order $n$, then the commutativity of $E/P^n$ is equivalent to the validity of the nth property of Pappus: If the cycles $U_1$, $U_2$, $U_3$ / $Z$ / $V_1$, $V_2$, $V_3$ of order $n$ in $D(G; E)$ are in Pappus order, if $W(i, j)=W(j, i)$ is for $i\neq j$ the cross-cut of $U_i+V_j$ and $U_j+V_i$, then*

$$W(1, 2) + W(2, 3) = W(2, 3) + W(3, 1) = W(3, 1) + W(1, 2).$$

**Proof.** Suppose first that the $n$th property of Pappus be satisfied by the cycles in $D(G; E)$, and suppose that $x$ and $y$ are two elements in $E$ such that neither $1+x$ nor $y$ is in $P$. There exist in $G$ three independent elements $z$, $u$, $v$ of order $n$; and the seven cycles

$$Z = zE,\quad \begin{matrix} U_1 = uE, & U_2 = (u - z)E, & U_3 = (u(1 + x) - zx)E, \\ V_1 = vE, & V_2 = (v - z)E, & V_3 = (vy + z)E \end{matrix}$$

are easily seen to be in Pappus order. Since consequently $W(2, 3)\leq W(2, 1)+W(1, 3)$, we infer from Lemma II.5.1 the existence of elements $r$, $s$ in $E$ such that

$$(v - z)y + (u(1 + x) - zx)(1 + y) = (u + v - z)r + (u(1 + x) - (vy + z)x)s;$$

and these elements $r$, $s$ must clearly satisfy the congruences

$$(1 + x)(1 + y) \equiv r + (1 + x)s,\ y\equiv r - yxs,\ y + x(1 + y) \equiv r + xs \bmod P^n.$$

Subtracting the third from the first congruence, we find $s\equiv 1$ mod $P^n$, and

hence it follows by subtracting the second from the third congruence that $xy \equiv yx \bmod P^n$.

If $1+x$ is in $P$, but $y$ is not in $P$, then $x$ is not in $P$; and $x=1+(x-1)$ together with the results already obtained imply that $(x-1)y\equiv y(x-1) \bmod P^n$ so that $xy\equiv yx \bmod P^n$, if at least one of the two elements $x$ and $y$ is not in $P$. If both are in $P$, then $1+x$ is not in $P$, so that $(1+x)y\equiv y(1+x) \bmod P^n$ and hence $xy\equiv yx \bmod P^n$; and this proves that $E/P^n$ is a commutative ring.

If conversely $E/P^n$ is a commutative ring, then any seven cycles of order $n$ in Pappus order may be assumed to be in the normal form (N) of Lemma II.5.1. Since $y(1+x)$ is not in $P$ and possesses therefore an inverse in $E$, and since we derive from the commutativity of multiplication the identity

$$\begin{aligned} &u(1 + x)(1 + y) + vy - z(y + x(1 + y)) - (u(1 + x) - vyx - zx) \\ &\qquad = u(1 + x)y + vy(1 + x) - zy(1 + x) = (u + v - z)y(1 + x), \end{aligned}$$

we find immediately that $W(1, 2)+W(2, 3)=W(2, 3)+W(3, 1)=W(3, 1)+W(1, 2)$, that is, the $n$th property of Pappus is satisfied in $D(G; E)$.

COROLLARY II.5.3. *If the nth property of Pappus is satisfied in $D(G; E)$ (and if $D(G; E)$ contains cycles of order n), then the $(n-1)$st property of Pappus is satisfied in $D(G; E)$.*

For $E/P^{n-1}$ is commutative, if $E/P^n$ is commutative.

COROLLARY II.5.4. *The nth property of Pappus is satisfied in $D(G; E)$ for every n if, and only if, E is commutative.*

For $E$ is commutative, if every $E/P^n$ is commutative, since 0 is the crosscut of the $P^n$.

6. **The ordinary primary abelian groups.** We assume throughout this section that $L$ is a primary ring of subgroups of the abelian group $G$ which either does not contain any cycles of order $n$ or at least three independent ones. It is the object of this section to find conditions on the (abstract) Dedekind set $L$ which assure that $L$ contains *every* subgroup of $G$.

If we denote by $E$ the ring of all the $L$-admissible endomorphisms of $G$, then it is a consequence of Theorem II.1.2 that $L$ is exactly the set $D(G; E)$ of all the $E$-admissible subgroups of $G$; and it is a consequence of Theorems II.2.2 and II.2.3 that $E$ is a primary ring of endomorphisms. Thus there exists in $E$ a uniquely determined greatest two-sided ideal $P$ different from $E$; and all the elements in $E$ that are not in $P$ possess inverses in $E$.

The integral multiples of the unit in $E$—which we denote by $0, \pm 1, \pm 2, \cdots$—form a subring $E_0$ of $E$. Clearly $E_0$ is part of the central of $E$; and the crosscut of $E_0$ and $P$ is either 0—in which case $E$ is said to be of characteristic[33])

[33]) The characteristic of the ring $E$ is exactly the characteristic of the field $E/P$ in the customary terminology.

0—or consists of all the multiples of a certain rational prime number $r$—in which case $E$ is said to be of characteristic[33] $r$.

If the set of all the subgroups of the abelian group $C$ is a cycle of order $n$, then $C$ is a cyclic group containing $c^n$ elements for $c$ a rational prime number. If the primary ring $L$ of subgroups of $G$ contains every subgroup of $G$, then the characteristic of $E$ is a rational prime number $r$, cycles of order $n$ in $L$ are cyclic groups of order $r^n$; and[34] $E$ is the ring of all the $r$-adic integers, provided the orders of the cycles in $L$ are not bounded; whereas $E$ is the ring of the rational integers modulo $r^m$, if $m$ is the maximum order of the cycles in $L$; in short: $G$ is a primary abelian group of characteristic $r$.

The six cycles $W, Z_1, Z_2; W_1, W_2, Z$ form *a complete triangle of order n* with *vertex* $W$ [and basis $Z_1+Z_2$], if

(a) $W, Z_1, Z_2$ are three independent cycles of order $n$ in $L$,

(b) $Z_1+Z_2=Z_2+Z=Z+Z_1$, $W_1+W_2=W_2+Z=Z+W_1$, $W+Z_i=Z_i+W_i$ $=W_i+W$.

A normal form for complete triangles is established by the following lemma.

Lemma II.6.1. *The six cycles $W, Z_1, Z_2; W_1, W_2, Z$ form a complete triangle of order n if, and only if, there exist in G three independent elements $w, z_1, z_2$ of order n such that $W=wE$, $Z_i=z_iE$, $W_i=(w+z_i)E$, $Z=(z_1-z_2)E$.*

**Proof.** The sufficiency of the condition is readily verified.—Thus we assume that the six cycles form a complete triangle. There exists an element $w$ such that $W=wE$. Since $W$ is part of $Z_i+W_i$, but of no proper part of this sum, there exist elements $z_i, w_i$ such that $w=w_i-z_i$, $W_i=w_iE$, $Z_i=z_iE$. Since $w, z_1, z_2$ are independent elements of order $n$, and since $Z$ is part of $Z_1+Z_2$ and $W_1+W_2$, but of no proper part of these sums, one may readily verify that $Z=(z_1-z_2)E$.

If the six cycles $W, Z_1, Z_2; W_1, W_2, Z$ form a complete triangle of order $n$ in $L$, then we define derived elements $W^{(j)}$, $W_1^{(j)}$ inductively for $j=-1, 0, 1, 2, \cdots$ by the following rules.

$$(-1)\qquad W=W^{(-1)},\qquad W_i=W_i^{(-1)}.$$

$$(j+1)\left\{\begin{array}{l} W^{(j+1)} \text{ is the cross-cut of } W \text{ and } W_1+W_1^{(j)};\\ V^{(j+1)} \text{ is the cross-cut of } W^{(j+1)}+Z \text{ and } W_1^{(j)}+W_2;\\ W_1^{(j+1)} \text{ is the cross-cut of } V^{(j+1)}+Z_2 \text{ and } Z_1+W^{(j+1)};\\ W_2^{(j+1)} \text{ is the cross-cut of } W_1^{(j+1)}+Z \text{ and } Z_2+W^{(j+1)}.\end{array}\right.$$

In particular we call $W^{(j)}$ the $j$th *derivative of the vertex of the complete triangle.*

Lemma II.6.2. *If the six cycles $W, Z_1, Z_2; W_1, W_2, Z$ form a complete triangle*

[34] Cf. R. Baer, American Journal of Mathematics, vol. 59 (1937), p. 110, Theorem 5.2.

*of order $n$ in $L$, and if $w$, $z_i$ are elements in $G$ such that $W=wE$, $Z_i=z_iE$, $W_i=(w+z_i)E$, $Z=(z_1-z_2)E$, then the derived elements satisfy:* $W^{(j)}=(wj)E$, $W_i^{(j)}=(z_i-wj)E$.

**Proof** (by induction). Our contention is obvious for $j=-1$ and thus we assume it to be true for $j$ in order to prove it for $j+1$. Using the fact that $w$, $z_1$, $z_2$ are three independent elements of order $n$, one verifies:

$W^{(j+1)}$ is the cross-cut of $wE$ and $(w+z_1)E+(z_1-wj)E$ and hence equals $w(j+1)E$;

$V^{(j+1)}$ is the cross-cut of $w(j+1)E+(z_1-z_2)E$ and $(z_1-wj)E+(w+z_2)E$ and hence equals $(z_1-z_2-w(j+1))E$;

$W_1^{(j+1)}$ is the cross-cut of $(z_1-z_2-w(j+1))E+z_2E$ and $z_1E+w(j+1)E$ and hence equals $(z_1-w(j+1))E$;

$W_2^{(j+1)}$ is the cross-cut of $(z_1-w(j+1))E+(z_1-z_2)E$ and $z_2E+w(j+1)E$ and hence equals $(z_2-w(j+1))E$;

and this completes the proof of the lemma.

THEOREM II.6.3. *If the primary ring $L$ of subgroups of the abelian group $G$ contains either no cycle of order $n$ or at least three independent ones, then the following conditions are necessary and sufficient for $L$ to contain every subgroup of $G$.*

(i) *If the subgroup $S$ in $L$ is the direct sum of two cycles of order 1 in $L$, then the number of subgroups in $L$ that are part of $S$ is $r+3$ for $r$ a rational prime number.*

(ii) *If $W$ is the vertex of some complete triangle of order $n$ in $L$, then there exists a superscript $j$ such that the $j$th derivative $W^{(j)}$ of $W$ is the subcycle of order $n-1$ of $W$.*

*If* (i) *and* (ii) *are satisfied, then $G$ is a primary abelian group of characteristic $r$.*

**Proof.** The necessity of condition (i) is an obvious consequence of the fact that $G$ is a primary abelian group of characteristic $r$, if $L$ contains every subgroup; and the necessity of (ii) is immediately derived from Lemmas II.6.1 and II.6.2.—Assume conversely that conditions (i) and (ii) are satisfied by $L$. If $L$ contains cycles of order $n$ different from 0, then $L$ contains three independent elements $w$, $z_1$, $z_2$ of order $n$. The six cycles $W=wE$, $Z_1=z_1E$, $Z_2=z_2E$; $W_1=(w+z_1)E$, $W_2=(w+z_2)E$, $Z=(z_1-z_2)E$ form a complete triangle of order $n$. Since the subcycle of order $n-1$ of $W$ is—by Theorem II.2.1—just $WP$, it follows from condition (ii) and from Lemma II.6.2 that there exists an integer $j$ such that $WP=wjE$; and this implies that $P=jE$ for $j$ in $E_0$.

It is a consequence of condition (i) that $E/P$ is a prime field of prime number characteristic $r$ so that $E/P^n$ consists of exactly $r^n$ elements, provided there exist—as we assume just now—cycles of order $n$ in $L$. If we denote by $P_0$

the cross-cut of $P$ and $E_0$, then it follows from $P=jE$ for $j$ in $E_0$, that $E_0/P_0^n$ consists of $r^n$ elements too so that $E/P^n$ and $E_0/P_0^n$ are essentially the same. If $G_n$ is the sum of all the cycles of an order not exceeding $n$ in $L$, then it follows from the fact that $L$ contains every $E$-admissible subgroup of $G$, that every subgroup of $G_n$ belongs to $L$, since every subgroup of $G_n$ is $E_0$-admissible, therefore $E/P^n$-admissible, therefore $E$-admissible. Since every element in $G$ is contained in some $G_n$, it follows finally that every subgroup of $G$ belongs to $L$ and that $G$ is a primary abelian group of characteristic $r$.

We add some remarks. If $S$ is any subset of the ring $L$ of subgroups of $G$, then we denote by $N(S)$ the *net* determined by $S$, that is, the smallest subset of $L$ which contains $S$ and which contains with any two elements their sum and their cross-cut. If, for example, $S$ consists of the six cycles of a complete triangle, then one may prove in a similar fashion as Theorem II.6.3 that $N(S)$ contains every part of the sum of the cycles of the triangle, provided $L$ contains every subgroup of $G$.

Applying Lemmas II.6.1 and II.6.2 one verifies immediately the following *characterization of the characteristic of $E$.*

Suppose that $W$ is the vertex of a complete triangle of order $n\neq 0$ in $L$. Then the characteristic of $E$ is 0 if, and only if, every derivative of the vertex $W$ is equal to $W$; and the characteristic of $E$ is the rational prime number $r$ if, and only if, the $r$th derivative $W^{(r)}$ of $W$ is a proper subcycle of $W$ (possibly 0).

## Part III. Construction of the underlying group

In the first part a class of Dedekind sets was determined which exhibited the salient features of both projective spaces and finite abelian groups. In the second part we introduced the primary abelian operator groups as those operator groups whose sets of admissible subgroups just met the requirements postulated in the first part. In this part we are going to prove that the (abstract) Dedekind sets with these properties may always be realized as the sets of admissible subgroups of a primary abelian operator group, provided they are "big enough." This last restriction is not surprising, considering the impossibility of obtaining a complete theory in the projective plane—there does not exist a planar proof of Desargues' theorem. The problem of determining the minimum number of parameters necessary for the validity of our theorem is still an open one; it is clear that it cannot be less than four and we prove that it is at most six.

The discussion in this part is based on the results of the two preceding parts. The method used is rather different from those customarily employed in dealing with similar problems of projective geometry, though an extension of Desargues' theorem is of central importance—and this part has some intrinsic interest apart from its applications. But in projective geometry it is customary to construct first the field of coordinates and then to introduce the linear subspaces by means of linear equations. Considering that the coordi-

nates are just the operators operating on the underlying group, this amounts to constructing the operators first and the group on which they operate only afterward. We invert this order of procedure, construct first the group with a set of distinguished subgroups representing the given Dedekind set and obtain operators as well as primarity as an application of theorems in Part II. Apart from this difference in the order of procedure one may say that the main difference consists of the true projectivity of our method and of the complete avoidance of affine means—the affine method consisting in first finding a representation for all the elements outside a certain distinguished hyperplane, a representation which is extended afterwards over the whole space, the projective method avoiding this preferential treatment of some element and thus finding representations of all the elements (as subgroups) at the same time.

1. **Collinearity.** The three elements $u$, $v$, $w$ in the Dedekind set $D$ are said to be *collinear*, if $u+v=v+w=w+u$.

LEMMA III.1.1. *If $u+w=w+v$, then $u$, $v$ and $w(u+v)$ are collinear.*

For it follows from Dedekind's law that

$$\begin{aligned} u+v=(u+v+w)(u+v) &= (u+w)(u+v) = u+w(u+v) \\ &= (v+w)(u+v) = v+w(u+v). \end{aligned}$$

LEMMA III.1.2. *If there exists an element $e$ such that $a$, $c$, $e$ and $b$, $d$, $e$ are triplets of collinear elements, then both $(a+b)(c+d)$, $c$, $d$ and $(a+b)(c+d)$, $a$, $b$ are triplets of collinear elements.*

This follows from Lemma III.1.1, since for example,

$$a+(c+d)=a+c+d=e+c+d=b+c+d=(c+d)+b.$$

2. **The theorem of Desargues.** The elements $w(i,j)=w(j,i)$ for $i\neq j$ are termed *connecting links of the two triplets $u(i)$ and $v(i)$*—$i=1, 2, 3$—of elements in the Dedekind set $D$, if $u(i)$, $w(i,j)$, $u(j)$ and $v(i)$, $w(i,j)$, $v(j)$ are two triplets of collinear elements for every $i\neq j$.

THEOREM III.2.1. *If the elements $w(i,j)$ are connecting links of the two triplets $u(i)$ and $v(i)$, and if $v(3)(u(1)+u(2)+u(3))=0$, then the $w(i,j)$ are collinear.*

**Proof.** If $i$, $j$, $h$ is any permutation of the three numbers 1, 2, 3, then it follows from the definition of connecting links that

$$\begin{aligned} u(1)+u(2)+u(3) &= w(i,j)+w(j,h)+u(3), \\ v(1)+v(2)+v(3) &= w(i,j)+w(j,h)+v(3); \end{aligned}$$

and hence it follows from the hypothesis and Dedekind's law that

$$w(i,j)+w(j,h)=w(i,j)+w(j,h)+v(3)(u(1)+u(2)+u(3))$$
$$=(w(i,j)+w(j,h)+v(3))(u(1)+u(2)+u(3))$$
$$=(v(1)+v(2)+v(3))(u(1)+u(2)+u(3)),$$

proving the collinearity of the connecting links $w(i, j)$.

The seven cycles $u(1)$, $u(2)$, $u(3)$ / $z$ / $v(1)$, $v(2)$, $v(3)$ are in *Desargues order*, if

(a) $z\neq 0$, $n(u(i))\leqq n(z)$, $n(v(i))\leqq n(z)$ for $i=1, 2, 3$,

(b) $u(i)$, $v(i)$, $z$ are collinear for $i=1, 2, 3$,

(c) $z(u(i)+u(j))(v(i)+v(j))=0$ for $i\neq j$([36]).

Since a cycle $c\neq 0$ contains one and only one subcycle $c^*$ of order 1—a notation we shall use throughout—condition (c) is equivalent to the following handier condition

(c′) $z(u(i)+u(j))=0$ or $z(v(i)+v(j))=0$ for $i\neq j$.

From these conditions we derive the following helpful statement

(d) $u(i)v(i)=0$; and if $z(u(i)+u(j))=0$, then $n(v(i))=n(z)$ and $v(i)(u(i)+u(j))=0$.

If $zu(i)=0$, then $z$, $(z+u(i))/u(i)=(u(i)+v(i))/u(i)$—by (b)—and $v(i)/(u(i)v(i))$ are isomorphic cycles; and hence it follows from (a) that $u(i)v(i)=0$ and $n(v(i))=n(z)$.—If $z(u(i)+u(j))=0$, though $v(i)(u(i)+u(j))\neq 0$, then $v(i)\neq 0$ and $v(i)^*\leqq u(i)+u(j)$. If $u(i)\neq 0$, then it follows from $u(i)v(i)=0$ and Theorem I.2.2 that $z^*\leqq u(i)^*+v(i)^*\leqq u(i)+u(j)$, contradicting our assumption. Thus $u(i)=0$; and this implies $v(i)^*=u(j)^*$ and $v(i)=z$—by (b)—so that $z^*=u(j)^*$, an impossibility which proves (d).

(e) The elements $w(i, j)=(u(i)+u(j))(v(i)+v(j))$ are the *only* connecting links of the two triples $u(i)$ and $v(i)$.

That the $w(i, j)$ are connecting links is an immediate consequence of Lemma III.1.2.—If the elements $x(i, j)$ are connecting links too, and if, for example, $z(u(i)+u(j))=0$, then $v(i)(u(i)+u(j))=0$ and hence

$$x(i,j)=x(i,j)+v(i)(u(i)+u(j))=(x(i,j)+v(i))(u(i)+u(j))$$
$$=(v(j)+v(i))(u(i)+u(j))=w(i,j)$$

by Dedekind's law.

(f) If $z+u(1)+u(2)+u(3)$ splits completely, then the connecting links $w(i, j)$ are cycles.

If $w(i, j)$ would not be a cycle, then it would follow from Theorem I.3.6 that $w(i, j)$ contains at least two different subcycles of order 1. Hence $u(i)\neq 0$, $u(i)^*\leqq v(i)+v(j)$ and $v(i)\neq 0$, $v(i)^*\leqq u(i)+u(j)$, contradicting (c′) and (d).

DESARGUES' PROPERTY. *If the seven cycles* $u(1)$, $u(2)$, $u(3)$ / $z$ / $v(1)$, $v(2)$, $v(3)$

([36]) It may be seen from trivial examples that this last condition is indispensable, though it seems to be fairly common to omit it.

*are in Desargues order, then their connecting links* $w(2, 3)$, $w(3, 1)$, $w(1, 2)$ *are collinear.*

We note that this property refers to all cycles in a certain Dedekind set.

THEOREM III.2.2. *If the element* $w$ *splits completely, is primary and contains at least five independent cycles of maximum order, then Desargues' property is satisfied by the cycles in the Dedekind set* $w/u$ *for every part* $u$ *of* $w$.

**Proof** ([30]). We show first that Desargues' property is satisfied by the subcycles of $w$; and we shall derive the general property from the special case.

Thus let us assume that the seven subcycles $z$, $u(i)$, $v(i)$ of $w$ are in Desargues order; and put $n(z)=m$. Since $z$, $u(i)$, $v(i)$ are collinear for every $i$, we find that $s=z+u(1)+u(2)+u(3)=z+v(1)+v(2)+v(3)$. Since there exist at least five independent subcycles of order $m$ of $w$, and since $s$ is a sum of four cycles only, it follows that there exists a subcycle $p$ of $w$ whose order is $m$ and which is independent of $s$.

Since $pz=0$, and since $p$ and $z$ are cycles of equal order $m$, it follows from Theorem I.3.7 and Lemma I.3.8 that there exists a subcycle $q$ of $p+z$ which is not part of any proper partial sum of $p+z$ and that $p+z$ is the direct sum of $p$ and $z$, of $z$ and $q$, of $q$ and $p$, $n(q)=m$.

We prove next that the elements $r(i)=(p+u(i))(q+v(i))$ are cycles.

For if this would not be true, then $r(i)$ would contain by Theorem I.3.6 two different subcycles of order 1; and since both $p+u(i)$ and $q+v(i)$ are sums of two cycles, this would imply that $q^*\leqq p+u(i)$, $p^*\leqq q+v(i)$ and $(p+u(i))^*=(q+v(i))^*$—where we denote by $r^*$ the sum of all the subcycles of order 1 of the element $r$, a notation which we shall use throughout. As a consequence of (c') we have $zu(i)=0$ or $zv(i)=0$. In the first case we obtain: $(p+u(i))^*=p^*+q^*=p^*+z^*=z^*+u(i)^*$, an inference contradicting $sp=0$; and in the same way we obtain a contradiction from $zv(i)=0$ so that the $r(i)$ have to be cycles.

Since $z$, $p$, $q$ and $z$, $u(i)$, $v(i)$ and $z$, $u(j)$, $v(j)$ are three triplets of collinear elements, it follows from Lemma III.1.2 that the elements $w(i,j)$, $r(j)$, $r(i)$ are connecting links of the two triplets $p$, $u(i)$, $u(j)$ and $q$, $v(i)$, $v(j)$—note that the elements $w(i, j)=(u(i)+u(j))(v(i)+v(j))$ are *the* connecting links of the triplets $u(i)$ and $v(i)$. From (c') it follows that not both $z(u(i)+u(j))$ and $z(v(i)+v(j))$ can be different from 0. If $z(u(i)+u(j))=0$, then it follows from $p(z+u(i)+u(j))=0$ and from Dedekind's law that $0=(p+z)(u(i)+u(j))$ $=(p+q)(u(i)+u(j))$; and this together with $pq=0$ implies $0=q(p+u(i)+u(j))$; and if $z(v(i)+v(j))=0$, then we prove likewise that $0=q(p+v(i)+v(j))$. Thus it follows in either case from Theorem III.2.1 that $w(i, j)$, $r(i)$ and $r(j)$ are collinear.

([30]) This is an adaptation of the proofs of Desargues' theorem as given in projective geometry; cf., for example, Veblen and Young, *Projective Geometry*, Boston, 1910, p. 41.

From $ps=0$, $r(i)\leqq p+u(i)$ and Dedekind's law we infer that

$$r(i)(u(1)+u(2)+u(3)) = r(i)(p+u(i))(u(1)+u(2)+u(3))$$
$$= r(i)(u(i)+p(u(1)+u(2)+u(3))) = r(i)u(i)$$

and likewise it follows that

$$r(i)(v(1)+v(2)+v(3)) = r(i)v(i).$$

Since $r(i)$ is a cycle, and since $u(i)v(i)=0$—by (d)—it is impossible that both $r(i)(u(1)+u(2)+u(3))$ and $r(i)(v(1)+v(2)+v(3))$ are different from 0. If $r(i)(u(1)+u(2)+u(3))=0$, then the connecting links $w(i,j)$ of the triplets $u(i)$ and $r(i)$ are collinear by Theorem III.2.1; and if $r(i)(v(1)+v(2)+v(3))=0$, then the connecting links $w(i,j)$ of the triplets $v(i)$ and $r(i)$ are collinear by Theorem III.2.1 so that the connecting links $w(i,j)$ of the triplets $u(i)$ and $v(i)$ are collinear in any case.

If $u$ is a part of $w$, and if $u(1), u(2), u(3) \,/\, z \,/\, v(1), v(2), v(3)$ are cycles in Desargues order in the Dedekind set $w/u$ of all the elements between $u$ and $w$ (whose null-element is $u$), then $s=z+u(1)+u(2)+u(3)=z+v(1)+v(2)+v(3)$. Since the parts of $w$ are sums of cycles by Theorem I.3.7 there exist subcycles $c, c(i)$ of $w$ (in $D$) such that $z=u+c$ and $u(i)=u+c(i)$. If $t=c+c(1)+c(2)+c(3)$, then $s=u+t$ and $s/u$ and $t/(tu)$ are isomorphic. Thus it suffices to prove that the cycles in the Dedekind set $t/(tu)$ satisfy Desargues' property. Since $t$ is a sum of four cycles, and since there exist at least five independent cycles of maximum order in $w$, there exists a subcycle $p$ of $w$ such that $pt=0$ and such that $n(p)=n(z/u)$. Now it is clear that the system $(p+t)/(tu)$ meets—as far as the subcycles of $t/(tu)$ are concerned—all the requirements we needed in the first part of the proof; and this completes the proof of our theorem.

3. **The vectors.** *A complete $n-m$-simplex $S=S_n^m$* consists of $n$ independent cycles $c(1), \cdots, c(n)$ of order $m$—the *vertices* of $S$—together with cycles $c(i,j)=c(j,i)$ for $i\neq j$—the *links* of $S$—subject to the following conditions.

(i) $c(i)$, $c(j)$ and $c(i,j)$ are collinear.

(ii) $c(i,j)$, $c(j,k)$ and $c(k,i)$ are collinear, if $i, j, k$ are three different integers.

Since $c(i)$ and $c(j)$ are for $i\neq j$ independent cycles of the same order $m$, and since $c(i,j)$ is a cycle, the order of $c(i,j)$ must be $m$ too and $c(i,j)c(k)=0$ for every $k$.

Lemma III.3.1. *If the cycles $c(1), \cdots, c(n)$ of order $m$ are independent, if $\sum_{i=1}^{n} c(i)$ splits completely and is primary, and if $4<n$, then the $c(i)$ are the vertices of some complete $n-m$-simplex.*

**Proof.** Since $c(1)+c(i)$ is for $1<i$ the direct sum of two cycles of equal order $m$, and is at the same time primary, it follows from Lemma I.3.8 that there exists a cycle $c(1,i)=c(i,1)$ of order $m$ such that $c(1)$, $c(i)$ and $c(1,i)$ are collinear.

If $i, j, k$ are three different integers which are all different from 1, then $c(1)(c(i)+c(j)+c(k))=0$ so that the seven cycles

$$c(1, i), c(1, j), c(1, k) \,/\, c(1) \,/\, c(i), c(j), c(k)$$

are in Desargues order; and their connecting links $c(j, k)$, $c(i, k)$, $c(i, j)$ are collinear by Theorem III.2.2 where $c(i, j)=(c(i)+c(j))(c(1, i)+c(1, j))$ is by (e), (f) of the preceding section a cycle (of order $m$); and hence a complete $n-m$-simplex is formed by the vertices $c(i)$ and the links $c(i, j)$.

If the cycles $c(i)$ are the vertices and the cycles $c(i, j)$ the links of a complete $n-m$-simplex $S$, then a *vector* $V$ *over* $S$ determines—and is itself determined by—the *cycle* $c(V)$ *generated by* $V$ and the *coordinates* $(i, V)$ for $i=1, \cdots, n$ subject to the following rules.

(a) $(i, V)=\infty$ *if, and only if,* $c(i)\, c(V)\neq 0$.

(b) $n(c(V))\leqq m$.

(c) *If* $c(i)\, c(V)=0$, *then* $(i, V)$ *is a cycle and* $c(i)$, $c(V)$, $(i, V)$ *are collinear.*

(d) *If* $c(i)\, c(V)=c(j)\, c(V)=0$, $i\neq j$, *then* $(i, V)$, $(j, V)$, $c(i, j)$ *are collinear.*

An immediate consequence of $n(c(V))\leqq n(c(i))$ and (c) is the following statement.

(e) $(i, V)\, c(V)=0$; $n((i, V))=m$.

THEOREM III.3.2. *If $S$ is an $n-m$-simplex with vertices $c(i)$ and links $c(i, j)$, if $4<n$, if $c$ is a cycle and $n(c)\leqq m$, if $c+\sum_{i=1}^{n}c(i)$ is primary and splits completely, if $c\, c(k)=0$, if $d$ is a cycle such that $c$, $d$, $c(k)$ are collinear, then there exists one and only one vector $V$ over $S$ such that $c=c(V)$ and $d=(k, V)$.*

**Proof.** We may assume without loss in generality that $k=1$.—Suppose first that $V$ and $U$ are vectors over $S$ such that $c(V)=c=c(U)$ and $(1, V)=d=(1, U)$. There exists at most one $i$ such that $c(1)(c+c(i))$ is not 0. If $c(1)(c+c(j))=0=c(1)(c+c(k))$ for $j\neq k$, then the seven cycles $c, c(j), c(k) \,/\, c(1) \,/\, d, c(1, j), c(1, k)$ are in Desargues order; and it follows from (e), (f) of the preceding section that their connecting links are uniquely determined and are $c(j, k)$, $(j, *)$, $(k, *)$; and hence it follows from (c) in the definition of a vector that $(j, *)=(j, U)=(j, V)$, $(k, *)=(k, V)=(k, U)$.—If finally $cc(i)=0$, but $c(1)(c+c(i))=0$, then it follows from similar reasoning that for some $j$ different from 1 and $i$

$$\begin{aligned}(i, V) &= (c + c(i))((j, V) + c(j, i))\\ &= (c + c(i))((j, U) + c(j, i)) = (i, U)\end{aligned}$$

so that $U=V$.

In order to prove the existence of a vector $V$ over $S$ meeting all the requirements we proceed in a similar fashion. There exists at most one $i\neq 1$ such that $c(1)(c+c(i))\neq 0$. If $j\neq 1$ and $c(1)(c+c(j))=0$, then we put $(j, V)=(c+c(j))(d+c(1, j))$. If $j\neq k$ and $c(1)(c+c(j))=0=c(1)(c+c(k))$, then the seven cycles $d, c(1, j), c(1, k) \,/\, c(1) \,/\, c, c(j), c(k)$ are in Desargues order and

their connecting links are $c(i, k)$, $(k, V)$, $(j, V)$. Hence it follows from (f) of the preceding section that $(j, V)$, $(k, V)$ are cycles, meeting the requirements (c) and (d) of the definition of a vector, as follows from (e) of the preceding section and from Theorem III.2.2. If we put $c=c(V)$, $d=(1, V)$, then the definition of the desired vector $V$ has been completed, if either $c(1)(c+c(i))=0$ for every $i\neq 1$ or $c(1)(c+c(i))\neq 0$ for an $i\neq 1$ such that $c(i)c\neq 0$, as we have to put $(i, V)=\infty$ in this case.

In order to complete the proof we assume now that $cc(n)=0$ and $c(1)(c+c(n))\neq 0$. But then $c(i)(c+c(n))=0$ for $1<i<n$; we put $(n, V)_i=(c+c(n))((i, V)+c(i, n))$. Then we may prove as before that $(n, V)_i$ is a cycle, that $c$, $c(n)$, $(n, V)_i$ as well as $(i, V)$, $c(i, n)$, $(n, V)_i$ are collinear. If furthermore, $1\leqq j<n$, $j\neq i$, then the seven cycles $c(i, n)$, $c(i, j)$, $(i, V)$ / $c(i)$ / $c(n)$, $c(j)$, $c=c(V)$ are in Desargues order; their connecting links $(j, V)$, $(n, V)_i$, $c(n, j)$ are collinear by Theorem III.2.2. But this implies $(n, V)_i\leqq(n, V)_j$ so that $(n, V)_i=(n, V)_j$ for reasons of symmetry. If we put now $(n, V)=(n, V)_2=\cdots=(n, V)_{n-1}$, then this completes again the definition of the required vector.

If $S$ is a complete $n-m$-simplex, $4<n$, if $c$ is a cycle such that $n(c)\leqq m$ and the sum of $c$ and of the vertices is primary and splits completely, then there exists a vertex $c(k)$ of $S$ such that $cc(k)=0$; and it follows from the primarity by Lemma I.3.8 that there exists a cycle $d$ such that $c$, $d$, $c(k)$ are collinear; and hence we have proved the following corollary to Theorem III.3.2.

*There exists a vector $V$ over $S$ which generates the cycle $c$.*

4. **Subtraction of vectors.** Throughout this section we shall assume that the element $g$ (in the Dedekind set $D$) is primary and splits completely, that $S$ is a complete $n-m$-simplex whose vertices $c(i)$ and whose links $c(i, j)$ are parts of $g$ and that $5<n$, though for some of the following results it would suffice to assume $4<n$.

LEMMA III.4.1. *If $A$ and $B$ are vectors over $S$ such that $c(A)$ and $c(B)$ are parts of $g$, and if $c(h)(c(A)+c(B))=0=(c(A)+c(B))c(k)$, then $(c(A)+c(B))((h, A)+(h, B))$ and $(c(A)+c(B))((k, A)+(k, B))$ are equal cycles and $c(A)$, $c(B)$, $(c(A)+c(B))((h, A)+(h, B))$ as well as $(h, A)$, $(h, B)$, $(c(A)+c(B))((h, A)+(h, B))$ are collinear triplets.*

**Proof.** We note first the existence of a $j$ such that $c(j)(c(A)+c(B)+c(h)+c(k))=0$, and that it suffices clearly to prove our statement for the couple $h, j$ instead of proving it for the couple $h, k$. Since $c(h)(c(A)+c(B))=0$, it follows that $c(j)+c(h)+c(A)+c(B)$ is the direct sum of $c(A)+c(B)$, $c(h)$, $c(j)$. Since $c(i)$, $c(X)$, $(i, X)$ are collinear whenever $c(i)c(X)=0$, it follows that the cycles $(h, A)$, $(h, B)$, $c(j, h)$ / $c(h)$ / $c(A)$, $c(B)$, $c(j)$ are in Desargues order; and their connecting links are the—by (e), (f) of §III. 2—uniquely determined cycles $(j, B)$, $(j, A)$, $((h, A)+(h, B))(c(A)+c(B))$ which are collinear by Theorem III.2.2.

On account of this lemma we *define*:

If $A$ and $B$ are vectors over $S$ such that $c(A)$ and $c(B)$ are parts of $g$, then $(A, B)=(c(A)+c(B))((i, A)+(i, B))$ for every $i$ such that $c(i)(c(A)+c(B))=0$.

We note that $(A, B)=(B, A)$ is a cycle such that both $c(A)$, $c(B)$, $(A, B)$ and $(i, A)$, $(i, B)$, $(A, B)$ are collinear triplets. This cycle $(A, B)$ shall serve later as the cycle generated by the difference of $A$ and $B$. For the definition of the coordinates of the difference vector we need two auxiliary functions.

LEMMA III.4.2. *If $A$ and $B$ are vectors over $S$ such that $c(A)$ and $c(B)$ are parts of $g$, then*

(a) $(c(A)+c(B))(c(i)+c(j))=0$ *and* $i\neq j$ *imply*

*that* $w(i/j; A, B)=(c(i, j)+(A, B))((i, A)+(j, B))$ *is a cycle of order* $m$,

*that* $(i, A)$, $(j, B)$, $w(i/j; A, B)$ *are collinear and their sum is the direct sum of any two of them,*

*that* $c(i, j)$, $(A, B)$, $w(i/j; A, B)$ *are collinear and that their sum is both the direct sum of* $c(i, j)$ *and* $(A, B)$ *and the direct sum of* $(A, B)$ *and* $w(i/j; A, B)$;

(b) $(c(A)+c(B))(c(i)+c(j))$ $(c(A)+c(B))(c(j)+c(h))=(c(A)+c(B))(c(h)+c(i))=0$ *for three different integers* $i, j, h$ *implies the collinearity of* $w(i/j; A, B)$, $w(h/j; A, B)$ *and* $c(i, h)$.

The formula defining $w(i/j; A, B)$ is meaningful if, and only if, $c(i)c(A)=c(B)c(j)=0$. But we shall consider this function only, if

$$(c(A)+c(B))(c(i)+c(j))=0.$$

**Proof.** If $(c(A)+c(B))(c(i)+c(j))=0$ and $i\neq j$, then it follows from property (d) of the vector definition and from Lemma III.4.1 that the two triplets $c(i, j)$, $(i, A)$, $(j, A)$ and $(A, B)$, $(j, B)$, $(j, A)$ are collinear; and hence it follows from Lemma III.1.2 that the triplets $(i, A)$, $(j, B)$, $w(i/j; A, B)$ and $c(i, j)$, $(A, B)$, $w(i/j; A, B)$ are collinear too.

Since it follows from our hypothesis and Lemma III.4.1 that $0=(c(A)+c(B))(c(i)+c(j))=((A, B)+c(A))(c(i)+c(i, j))$, and since $c(A)(i, A)=0$ by (e) of the preceding section, it follows from Lemma I.1.1 that

$$\begin{aligned}(i, A)w(i/j; A, B) &= (i, A)(c(i) + c(A))(c(i, j) + (A, B))w(i/j; A, B)\\ &= (i, A)(c(i)c(i, j) + c(A)(A, B))w(i/j; A, B)\\ &= (i, A)c(A)(A, B)w(i/j; A, B) = 0;\end{aligned}$$

and likewise it follows that

$$\begin{aligned}(i, A)(j, B) &= (i, A)(c(i) + c(A))(c(j) + c(B))(j, B)\\ &= (i, A)(c(i)c(j) + c(A)c(B))(j, B)\\ &= (i, A)c(A)c(B)(j, B) = 0.\end{aligned}$$

The sum of the collinear triplet $(i, A)$, $(j, B)$, $w(i/j; A, B)$ is therefore the direct sum of the two cycles $(i, A)$ and $(j, B)$ of order $m$; and $(i, A)\, w(i/j; A, B) = 0$ implies therefore that $w(i/j; A, B)$ cannot contain more than one subcycle of order 1. Thus it follows from Theorem I.3.7 that $w(i/j; A, B)$ is a cycle; and the remainder of (a) is readily proved.

In order to prove (b) we assume first in addition to the special hypotheses of (b) that $0 = c(A)(c(i)+c(j)+c(h))$. Then it follows from Lemma I.1.1 and $(c(A)+(A, B))(c(j)+c(i, j)) = 0$ that

$$\begin{aligned}(j, A)(c(i, j) + (A, B)) &= (j, A)(c(j) + c(A))(c(i, j) + (A, B)) \\ &= (j, A)(c(j)c(i, j) + c(A)(A, B)) \\ &= (j, A)c(A)(A, B) = 0\end{aligned}$$

and likewise that $(j, A)(c(h, j)+(A, B)) = 0$; and it follows from Dedekind's law and $c(j)(c(i, j)+c(j, h)) = 0$ that

$$\begin{aligned}&(j, A)(c(i, j) + c(j, h)) \\ &\quad = (j, A)(c(j) + c(A))(c(i) + c(j) + c(h))(c(i, j) + c(j, h)) \\ &\quad = (j, A)(c(j) + c(A)(c(i) + c(j) + c(h)))(c(i, j) + c(j, h)) \\ &\quad = (j, A)c(j)(c(i, j) + c(j, h)) = 0.\end{aligned}$$

Hence we have shown that the seven cycles $c(i,j), c(h,j), (A, B) / (j, A) / (i, A), (h, A), (j, B)$ are in Desargues order; and it follows from part (a) of this lemma that their uniquely determined connecting links are the cycles $w(h/j; A, B)$, $w(i/j; A, B)$, $c(i, h)$ and that their collinearitv is a consequence of Theorem III.2.2.

To derive the general case of (b) from the special case already proved we note that $c(A)+c(B)+c(i)+c(j)+c(h)$ is a sum of five cycles; and hence it follows from $5<n$ and from Corollary I.3.4 that there exists an integer $k$, different from $i, j, h$, such that

$$c(k)(c(A) + c(B) + c(i) + c(j) + c(h)) = 0.$$

Thus it follows from $(c(A)+c(B))(c(i)+c(j)) = 0$ that

$$(c(A) + c(B))(c(i) + c(j) + c(k)) = 0;$$

and likewise that

$$\begin{aligned}(c(A) + c(B))(c(j) + c(h) + c(k)) &= 0, \\ (c(A) + c(B))(c(h) + c(i) + c(k)) &= 0;\end{aligned}$$

and hence it follows from what we have shown in the preceding paragraph that $w(i/j; A, B)$, $w(k/j; A, B)$, $c(i, k)$ and $w(k/j; A, B)$, $w(h/j; A, B)$, $c(k, h)$ are collinear triplets.

Using (a) and Lemma I.1.1 it follows that

$$(k, A)(w(k/j; A, B) + c(k, h))$$
$$= (k, A)(c(k) + c(A))((A, B) + c(k, j) + c(k, h))(w(k/j; A, B) + c(k, h))$$
$$= (k, A)(c(A)(A, B) + c(k)(c(k, j) + c(k, h)))(w(k/j; A, B) + c(k, h))$$
$$= (k, A)c(A)(A, B)(w(k/j; A, B) + c(k, h)) = 0$$

and likewise that

$$(k, A)(w(k/j; A, B) + c(k, i)) = 0;$$

and since $(k, A)(c(k, h)+c(k, i))=0$ may be shown as before, it follows that the seven cycles $w(k/j; A, B)$, $c(k, h)$, $c(k, i)$ / $(k, A)$ / $(j, B)$, $(h, A)$, $(i, A)$ are in Desargues order; and hence it follows from Theorem III.2.2 that their uniquely determined connecting links $c(i, h)$, $w(i/j; A, B)$, $w(h/j; A, B)$ are collinear, completing the proof of (b).

The following two remarks will simplify the handling of the function $w(i/j; A, B)$. It is a consequence of its defining equation and of $(A, B)=(B, A)$ that

$$w(i/j; A, B) = w(j/i; B, A);$$

and hence it follows under the hypotheses of Lemma III.4.2, (b) that $w(i/j; A, B)$, $w(i/h; A, B)$, $c(j, h)$ *are collinear.*

LEMMA III.4.3. *If $A$ and $B$ are vectors over $S$ such that $c(A)$ and $c(B)$ are parts of $g$, then*

(a) $(c(A)+c(B))(c(i)+c(j))=0$ *and* $i\neq j$ *imply that* $z(i/j; A, B) = (w(i/j; A, B)+c(j))(c(i)+(A, B))$ *is a cycle of order* $m$,

*that* $z(i/j; A, B)$, $w(i/j; A, B)$, $c(j)$ *are collinear and their sum is the direct sum of any two of them,*

*that* $z(i/j; A, B)$, $(A, B)$, $c(i)$ *are collinear and their sum is both the direct sum of* $z(i/j; A, B)$ *and* $(A, B)$ *and the direct sum of* $(A, B)$ *and* $c(i)$,

*that* $z(i/j; A, B)$, $(i, A)$, $c(B)$ *are collinear;*

(b) $(c(A)+c(B))(c(i)+c(j))=(c(A)+c(B))(c(i)+c(h))=0$ *for three different integers* $i, j, h$ *imply* $z(i/j; A, B)=z(i/h; A, B)$.

**Proof.** Since it follows from Lemma III.4.2 and the definition of a complete simplex that the triplets $w(i/j; A, B)$, $(A, B)$, $c(i, j)$ and $c(j)$, $c(i)$, $c(i, j)$ are collinear, we may infer from Lemma III.1.2 that the triplets $z(i/j; A, B)$, $w(i/j; A, B)$, $c(j)$ and $z(i/j; A, B)$, $c(i)$, $(A, B)$ are collinear. Since $c(A)+c(B)+c(i)+c(j)$ is the direct sum of $c(A)+c(B)$, $c(i)$ and $c(j)$, it follows from $(A, B)\leqq c(A)+c(B)$ and $z(i/j; A, B)\leqq c(i)+(A, B)$ that $z(i/j; A, B)c(j)=0$; and similarly we see that $w(i/j; A, B)c(j)=0$. From these facts one derives as usual all the statements of (a) except the last one. This last statement is a consequence of Theorem III.2.1, since $z(i/j; A, B)$, $(i, A)$, $c(B)$ are connecting links of the two triplets $c(A)$, $(A, B)$, $c(i)$ and $(j, B)$, $c(j)$, $w(i/j; A, B)$, and since $c(j)(c(A)+(A, B)+c(i))=0$, as has been pointed out before.

In order to prove (b) we assume first that—in addition to the other hypotheses—$(c(A)+c(B))(c(i)+c(j)+c(h))=0$. Then $c(i, h)((A, B)+c(j, i))=0=c(i, h)((A, B)+c(i))$; and thus the seven cycles $c(j, i)$, $c(i)$, $(A, B)$ / $c(i, h)$/ $c(j, h)$, $c(h)$, $w(i/h; A, B)$ are in Desargues order, as follows from Lemma III.4.2 and the properties of a complete simplex. Thus their connecting links are uniquely determined; they are—by Lemma III.4.2—the cycles $z(i/h; A, B)$, $w(i/j; A, B)$, $c(j)$; and the collinearity of these cycles is a consequence of Theorem III.2.2. Thus

$$z(i/h; A, B) \leqq (w(i/j; A, B) + c(j))(c(i) + (A, B)) = z(i/j; A, B);$$

and this inequality implies equality, since every $z(\cdots)$ has been shown to be a cycle of order $m$.

To derive the general case of (b) from the special case we have proved just now, we note first that there exists an integer $k$ such that $c(k)(c(i)+c(j)+c(h)+c(A)+c(B))=0$, since $5<n$, since $c(i)+c(j)+c(h)+c(A)+c(B)$ is a sum of five cycles and therefore by Corollary I.3.4 a direct sum of at most five cycles. Hence it follows from our hypothesis that $(c(A)+c(B))(c(i)+c(j)+c(k))=0=(c(A)+c(B))(c(i)+c(h)+c(k))$; and from what has been shown already it follows that

$$z(i/j; A, B) = z(i/k; A, B) = z(i/h; A, B),$$

as was to be shown.

If $A$ and $B$ are vectors over $S$ such that $c(A)$ and $c(B)$ are parts of $g$, then there exist integers $i$ such that $c(i)(c(A)+c(B))=0$; and to every $i$ satisfying this condition there exist integers $j\neq i$ such that $(c(i)+c(j))(c(A)+c(B))=0$. It is a consequence of Lemma III.4.3, (b) that the cycle $z(i/j; A, B)$ is independent of the choice of $j$; and thus the following *definition* is well determined.

Definition. *If $c(i)(c(A)+c(B))=0$, then $z(i; A, B)=z(i/j; A, B)$ for every $j\neq i$ such that $(c(i)+c(j))(c(A)+c(B))=0$.*

Lemma III.4.4. *If $A$ and $B$ are vectors over $S$ such that $c(A)+c(B)\leqq g$; and if $c(i)(c(A)+c(B))=c(j)(c(A)+c(B))=0$ for $i\neq j$, then $z(i; A, B)$, $z(j; A, B)$ and $c(i, j)$ are collinear.*

**Proof.** From $5<n$ we infer as usual the existence of an integer $h$ such that $c(h)(c(i)+c(j)+c(A)+c(B))=0$. From the hypothesis of this lemma it follows now that $(c(i)+c(h))(c(A)+c(B))=0=(c(j)+c(h))(c(A)+c(B))$ so that $z(i; A, B)=z(i/h; A, B)$, $z(j; A, B)=z(j/h; A, B)$. The three cycles $z(i; A, B)$, $z(j; A, B)$, $c(i, j)$ are therefore connecting links of the two triplets $c(j)$, $c(i)$, $(A, B)$ and $w(j/h; A, B)$, $w(i/h; A, B)$, $c(h)$, as follows from Lemma III.4.2, (b) and Lemma III.4.3, (a); and the collinearity of these connecting links is a consequence of Theorem III.2.1, since $c(h)(c(i)+c(j)+(A, B))=0$.

If $A$ and $B$ are any two vectors over $S$ such that $c(A)+c(B)\leqq g$, then there exist integers $i\neq j$ such that $(c(A)+c(B))(c(i)+c(j))=0$. Thus $(A, B)$ is a

well determined cycle (contained in $c(A)+c(B)$ so that $n((A, B))\leqq m$ by Theorem I.2.1) by Lemma III.4.1 and the cycles $z(i; A, B)$ and $z(j; A, B)$ are well determined by Lemma III.4.3, (b). It follows finally from Lemma III.4.4 and Theorem III.3.2 that there exists one and only one vector $D$ over $S$ such that $c(D)=(A, B)$ and such that $(i, D)=z(i; A, B)$ for every $i$ such that $c(i)(c(A)+c(B))=0$; and this vector $D$ over $S$ which is uniquely determined by $A$ and $B$ shall be termed *the difference $A-B$ of $A$ and $B$.*

There exists one and only one vector 0 over $S$ satisfying $c(0)=0$ and $(i, 0)=c(i)$ for every $i$; and one verifies readily that 0 is the only vector $V$ over $S$ such that $c(V)=0$.

LEMMA III.4.5. *If $A$ is a vector over $S$ such that $c(A)\leqq g$, then*
(i) $A-A=0$,
(ii) $A=A-0$,
(iii) $A=0-(0-A)$.

**Proof.** There exist integers $i\neq j$ such that $c(A)(c(i)+c(j))=0$. Thus it follows from Lemma III.4.1 that $c(A-A)=(A, A)=c(A)(i, A)=0$ by (e) of §3 of this part and this implies $A-A=0$ by Lemma III.4.3. It follows furthermore from Lemma III.4.1 and (c) of §3 that

$$c(A-0)=(A, 0)=(c(A)+0)((i, A)+c(i))=c(A)((i, A)+c(i))=c(A);$$

and hence it follows from Lemma III.4.3 that

$$\begin{aligned}(i, A-0)&=z(i; A, 0)=z(i/j; A, 0)=(c(i)+c(A))(c(j)+w(i/j; A, 0))\\&=(c(i)+c(A))(c(j)+(c(i, j)+c(A))((i, A)+c(j)))\\&=(c(i)+c(A))(c(j)+c(i, j)+c(A))((i, A)+c(j))\\&=(c(i)+c(A))((i, A)+c(j))=(i, A)\quad\text{or}\quad A-0=A.\end{aligned}$$

To prove (iii) we note that $c(0-A)=(0, A)=(A, 0)=c(A-0)=c(A)$ and that therefore $w(i/j; 0, A)=(c(i, j)+c(A))(c(i)+(j, A))$, $c(A)\leqq c(i, j)+c(A)$ $=c(i, j)+w(i/j; 0, A)$ by Lemma III.4.2, (a) and $(i, 0-A)=z(i/j; 0, A)$ $=(c(i)+c(A))(c(j)+w(i/j; 0, A))$ so that

$$\begin{aligned}c(0-(0-A))&=(0, 0-A)=(0+c(A))(c(i)+(i, 0-A))\\&=c(A)(c(i)+(c(i)+c(A))(c(j)+w(i/j; 0, A)))\\&=c(A)(c(i)+c(j)+w(i/j; 0, A))(c(i)+c(A))=c(A)\end{aligned}$$

by Dedekind's law and Lemma III.4.1; and consequently it follows from Lemmas III.4.1, III.4.3 and III.4.4 that

$$\begin{aligned}w(i/j; 0, 0-A)&=(c(i, j)+c(A))(c(i)+(j, 0-A))\\&=(c(i, j)+c(A))(c(i)+z(j/i; 0, A))\\&=(c(i, j)+c(A))(c(i)+w(j/i; 0, A))\\&=(c(i, j)+(A, 0))(c(i)+w(i/j; A, 0))=w(i/j; A, 0)\end{aligned}$$

and

$$(i, 0 - (0 - A)) = (c(i) + c(A))(c(j) + w(i/j; 0, 0 - A))$$
$$= (c(i) + c(A))(c(j) + w(i/j; A, 0)) = z(i/j; A, 0) = (i, A)$$

by (ii); and hence $0-(0-A)=A$.

LEMMA III.4.6. *If $A, B, C$ are vectors over $S$ such that $s=c(A)+c(B)+c(C)$ is contained in $g$, then $(A-B)-C=(A-C)-B$.*

The proof of this associative law of subtraction will be effected in a number of steps.

(1) *$c(A-B)$, $c(B-C)$, $c(C-A)$ are collinear.*

There exists an integer $i$ such that $sc(i)=0$. Then the seven cycles $(i, A)$, $(i, B)$, $(i, C)$ / $c(i)$ / $c(A)$, $c(B)$, $c(C)$ are in Desargues order and their uniquely determined connecting links $c(B-C)$, $c(C-A)$, $c(A-B)$—by Lemma III.4.1—are collinear by Theorem III.2.2.

(2) *If $(c(i)+c(j))s=0$ for $i\neq j$, then the two triplets $c(B-C)$, $w(i/j; A, C)$, $w(i/j; A, B)$ and $c(A-B)$, $w(i/j; A, C)$, $w(i/j; B, C)$ are collinear.*

The seven cycles $c(A-B)$, $c(A-C)$, $c(i, j)$ / $(j, A)$ / $(j, B)$, $(j, C)$, $(i, A)$ are in Desargues order by Lemma III.4.1 and the definition of vectors, since

$$\begin{aligned}(j, A)(c(A-B) + c(A-C) + c(i, j)) &\leqq (j, A)(c(j) + c(A))(c(i, j) + s)\\ &\leqq (j, A)(c(A) + c(j)(s + c(i, j)))\\ &= (j, A)c(A) = 0;\end{aligned}$$

their connecting links are by (1) and Lemma III.4.2 the cycles $w(i/j; A, C)$, $w(i/j; A, B)$, $c(B-C)$ and their collinearity is a consequence of Theorem III.2.2.—The collinearity of the second triplet is immediately inferred from the collinearity of the first triplet, if one remembers that $w(i/j; X, Y) = w(j/i; Y, X)$.

(3) *If $c(i)s=0$, then the two triplets $(i, A-B)$, $(i, A-C)$, $c(B-C)$ and $(i, A-C)$, $(i, B-C)$, $c(A-B)$ are collinear.*

There exists some $j\neq i$ such that $s(c(i)+c(j))=0$. Then the cycles of the first triplet are connecting links of the two triplets $c(A-C)$, $c(A-B)$, $c(i)$ and $w(i/j; A, C)$, $w(i/j; A, B)$, $c(j)$, as follows from (1), (2), Lemma III.4.3; and thus the collinearity of the first triplet is a consequence of Theorem III.2.1, since $c(j)(c(i)+c(A-C)+c(A-B))\leqq c(j)(c(i)+s)=0$.—The cycles of the second triplet are likewise connecting links of the triplets $(i, B)$, $(i, A)$, $c(C)$ and $w(i/j; B, C)$, $w(i/j; A, C)$, $c(j)$; and the collinearity of the second triplet follows from Theorem III.2.1, since $c(j)(c(C)+(i, A)+(i, B)) \leqq c(j)(c(i)+s)=0$.

Since $5<n$, there exist three different integers $i, j, h$ such that $(c(i)+c(j)+c(h))s=0$; and these three integers shall be kept fixed throughout the remainder of the proof.

(4) *$w(i/j; B, C), w(j/h; C, A), w(h/i; A, B)$ are collinear.*

One verifies readily that $c(A-B)(j, C)=(c(A-B)+(j, C))(i, B)=(c(A-B)+(j, C)+(i, B))(h, A)=0$. It follows from (2) and Lemma III.4.2 that $w(i/j; B, C), w(j/h; C, A), w(h/i; A, B)$ are connecting links of the two triplets $(h, A), (i, B), (j, C)$ and $c(h, i), c(A-B), w(i/j; A, C)$; and (4) is now a consequence of Theorem III.2.1, since the above equations imply $c(A-B)((h, A)+(i, B)+(j, C))=0$.

We introduce now three auxiliary vectors. If $X$ is any of the vectors $A, B, C$ under consideration, then $X_h$ is the—by Theorem III.3.2 uniquely determined—vector satisfying $c(X_h)=(h, 0-X)$, $(i, X_h)=w(h/i; 0, X)$. It is an immediate consequence of Lemmas III.4.2 and III.4.3 that $(j, X_h)=w(h/j; 0, X)$. We are not able to say anything concerning the $h$-coordinate of this vector $X_h$.

We note furthermore that $(c(i)+c(j))(c(A_h)+c(B_h)+c(C_h))=(c(i)+c(j))((h, 0-A)+(h, 0-B)+(h, 0-C))\leqq(c(i)+c(j))(c(h)+s)=0$ so that the statements (2) and (3) may be applied upon $A_h, B_h, C_h$ too.

(5) *If $X$ and $Y$ are two of the vectors $A, B, C$, then $X-Y=X_h-Y_h$.*

It is a consequence of $s(c(i)+c(j)+c(h))=0$ and of (2), (3) that the two triplets $c(X-Y), (h, 0-X), (h, 0-Y)$ and $c(X-Y), w(h/i; 0, X), w(h/i; 0, Y)$ are collinear triplets. Since furthermore

$$
\begin{aligned}
&(h, 0-X)(w(h/i; 0, X)+w(h/i; 0, Y))\\
&\qquad\qquad\leqq (h, 0-X)(c(h)+c(X))(c(h, i)+c(X)+c(Y))\\
&\qquad\qquad\leqq (h, 0-X)c(X)=0,
\end{aligned}
$$

it follows from Dedekind's law and the definition of the cycle generated by the difference of vectors (cf. Lemma III.4.1) that

$$
\begin{aligned}
c(X_h-Y_h) &= ((h, 0-X)+(h, 0-Y))(w(h/i; 0, Y)+w(h/i; 0, X))\\
&= c(X-Y)+(h, 0-X)(w(h/i; 0, Y)+w(h/i; 0, X))\\
&= c(X-Y).
\end{aligned}
$$

It is a consequence of (4) that the cycles $w(i/j; X, Y), w(j/h; Y, 0), w(h/i; 0, X)$ are collinear; and it is a consequence of Lemma III.4.2 that the cycles $c(i, j), c(X-Y), w(i/j; X, Y)$ are collinear. Thus it follows that $w(i/j; X, Y)\leqq(c(i, j)+c(X-Y))(w(h/i; 0, X)+w(h/j; 0, Y))=(c(i, j)+c(X_h-Y_h))((i, X_h)+(i, Y_h))=w(i/j; X_h, Y_h)$; and this inequality implies equality, since the cycles $w(i/j; \cdots)$ are of order $m$. From the equalities thus obtained, it follows that $z(i/j; X, Y)=z(i/j; X_h, Y_h)$ (cf. Lemma III.4.3); and hence $(i, X-Y)=(i, X_h-Y_h)$ and $X-Y=X_h-Y_h$ is now a consequence of Theorem III.3.2.

(6) *$(i, A-C), (j, A-B), w(i/j; B, C)$ are collinear.*

It follows from (2) that $w(i/j; B_h, 0), c(C_h)=c(C_h-0)$—by Lemma III.4.5, (ii)—and $w(i/j; B_h, C_h)$ are collinear; and from (3) that $(i, A_h-B_h), c(B_h-C_h)$,

$(i, A_h-C_h)$ are collinear. The three cycles $(i, A_h)$, $(j, A_h-B_h)$, $w(i/j; B_h, 0)$ are collinear, since they are—by Lemmas III.4.1 to III.4.3—connecting links of the two triplets $c(B_h)$, $c(i, j)$, $(j, A_h)$ and $c(j)$, $(i, B_h)$, $c(A_h-B_h)$, and since Theorem III.2.1 may be applied as a consequence of

$$\begin{aligned} c(B_h)(c(j) + (i, B_h) + c(A_h - B_h)) &\\ &= (h, 0 - B)(c(j) + w(h/i; 0, B) + c(A - B)) \\ &\leqq (h, 0 - B)(c(h) + c(B))(c(j) + c(h, i) + c(A) + c(B)) \\ &= (h, 0 - B)c(B) = 0. \end{aligned}$$

Since $c(B_h)=(h, 0-B)$ is a cycle of order $m$, and since $c(B_h)(c(i, j) + c(B_h - C_h) + (i, A_h - B_h)) = (h, 0 - B)(c(i,j) + c(B - C) + (i, A - B)) \leqq (h, 0-B)(c(i)+c(j)+c(A)+c(B)+c(C))=0$, it follows that the seven cycles $w(i/j; B_h, 0)$, $c(C_h)$, $(i, A_h)$ / $c(B_h)$ / $c(i, j)$, $c(B_h-C_h)$, $(i, A_h-B_h)$ are in Desargues order, that $(i, A_h-C_h)$, $(j, A_h-B_h)$, $w(i/j; B_h, C_h)$ are their uniquely determined connecting links; and (6) is now a consequence of Theorem III.2.2 and of (5).

(7) $c((A-B)-C)=c((A-C)-B)$.

The three cycles $c((A-B)-C)$, $(i, A-C)$, $(i, B)$ are collinear, since they are—by (b)—connecting links of the two triplets $(i, A)$, $c(A-B)$, $c(C)$ and $w(i/j; B, C)$, $(j, C)$, $(j, A-B)$, and since Theorem III.2.1 may be applied as a consequence of $(j, C)((i, A)+c(A-B)+c(C)) \leqq (j, C)(c(i)+s))=0$.

$c(i)s=0$ implies that the seven cycles $(i, A)$, $(i, A-B)$, $(i, C)$ / $c(i)$ / $c(A)$, $c(A-B)$, $c(C)$ are in Desargues order; and it follows from Theorem III.2.2 that their connecting links $c((A-B)-C)$, $c(A-C)$, $c(B)$ are collinear. Consequently we find that

$$c((A - B) - C) \leqq (c(A - C) + c(B))((i, A - C) + (i, B)) = c((A - C) - B)$$

and the symmetry of our hypotheses on $B$ and $C$ implies the opposite inequality, proving the desired equation (7).

(8) $c(h)=c(A-A_h)$.

Since $c(h) \leqq ((h, 0-A)+c(A))(w(h/i; 0, A)+(i, A))=c(A_h-A)$ as a consequence of Lemmas III.4.1 to III.4.4, the equation (8) may be inferred from the fact that $n(c(A-A_h)) \leqq m=n(c(h))$.

(9) *$c(h)$, $c(A_h-X)$, $c(A-X)$ and $c(h)$, $(k, A_h-X)$, $(k, A-X)$ are collinear triplets for $X=B$, $C$ and $k=i, j$.*

This is an immediate consequence of (8), (4) and (1)

(10) *$c(h)$, $c((A_h-B)-C)$, $c((A-B)-C)$ are collinear.*

For they are connecting links of the seven cycles $(i, C)$, $(i, A_h-B)$, $(i, A-B)$ / $c(i)$ / $c(C)$, $c(A_h-B)$, $c(A-B)$ which are in Desargues order, since $c(i)(c(A-B)+c(A_h-B)+c(C)) \leqq c(i)(c(h)+s)=0$, and since therefore Theorem III.2.2 may be applied.

(11) *$c(h)$, $w(i/j; A_h-B, C)$, $w(i/j; A-B, C)$ are collinear.*

Since $(i, C)$ is a cycle of order $m$, since $(i, C)(c(i, j)+c((A-B)-C)+c((A_h-B)-C)) \leqq (i, C)(c(i,j)+c(h)+s)=0$, it follows that the seven cycles $c(i, j)$, $c((A-B)-C)$, $c((A_h-B)-C)$ / $(i, C)$ / $(j, C)$, $(i, A-B)$, $(i, B_h-B)$ are in Desargues order so that their connecting links—by (10) and (9)—$c(h)$, $w(i/j; A_h-B, C)$, $w(i/j; A-B, C)$ are collinear as a consequence of Theorem III.2.2.

(12) $(j, B)$, $(i, A-C)$, $w(i/j; A-B, C)$ *are collinear.*

The seven cycles $c(B-C)$, $(j, C)$, $(i, A_h-B)$ / $(i, C)$ / $(i, B)$, $c(i, j)$, $c((A_h-B)-C)$ are in Desargues order, since $(i, C)$ is a cycle of order $m$, and since $(i, C)(c(B-C)+(j, C)+(i, A_h-B)) \leqq (i, C)(s+c(j)+(i, A_h)) \leqq (i, C)(s+c(j)+w(h/i; 0, A))=(i, C)(s+c(j)+c(h, i))=0$. Since the cycles $w(i/j; A_h-B, C)$, $(i, A_h-C)$, $(j, B)$ are by (3) and (7) their connecting links, it follows from Theorem III.2.2 that they are collinear. Using this fact and (9), (11) we find that

$$\begin{aligned} c(h) + w(i/j; A - B, C) + (j, B) &= c(h) + w(i/j; A_h - B, C) + (j, B) \\ &= c(h) + (j, B) + (i, A_h - C) \\ &= c(h) + (i, A - C) + w(i/j; A - B, C) \\ &= c(h) + (i, A_h - C) + w(i/j; A_h - B, C). \end{aligned}$$

Since $c(h)((j, B)+(i, A-C)+w(i/j; A-B, C)) \leqq c(h)(s+c(j)+c(i))=0$, these identities imply that

$$\begin{aligned} &(j, B) + (i, A - C) + w(i/j; A - B, C) \\ &\quad = (j, B) + (i, A - C) \\ &\quad = (i, A - C) + w(i/j; A - B, C) = (j, B) + w(i/j; A - B, C), \end{aligned}$$

as was to be shown.

It follows now from (12) and (7) and Lemmas III.4.1 to III.4.4 that $w(i/j; A-B, C) \leqq ((j,B)+(i,A-C))(c(i,j)+c((A-C)-B))=w(i/j; A-C, B)$; thus $w(i/j; A-B, C)=w(i/j; A-C, B)$, since they are both cycles of order $m$. But now it follows from (7) that $z(i/j; A-B, C)=z(i/j; A-C, B)$ and that consequently $z(i; A-B, C)=z(i; A-C, B)$ or $(i, (A-B)-C)=(i, (A-C)-B)$; and $(A-B)-C=(A-C)-B$ is now a consequence of (7) and Theorem III.3.2.

Lemma III.4.7. *If $t \leqq g$, then the set $(S; t)$ of all the vectors $V$ over $S$ which satisfy $c(V) \leqq t$ is an abelian group (under the definition of subtraction introduced in connection with Lemma* III.4.4).

**Proof.** If the vectors $A$ and $B$ over $S$ belong to $(S; t)$, then it follows from Lemmas III.4.4 and III.4.1 that $A-B$ is a uniquely determined vector over $S$ which satisfies: $c(A-B) \leqq c(A)+c(B) \leqq t$; and thus $(S; t)$ contains with any two vectors over $S$ their uniquely determined difference. It is a consequence

of Lemmas III.4.5 and III.4.6 that this subtraction in $(S; t)$ satisfies furthermore the following rules.

(A) There exists one and only one element 0 in $(S; t)$ satisfying $0=A-A$, $A=A-0=0-(0-A)$ for every $A$ in $(S; t)$.

(B) If $A, B, C$ are elements in $(S; t)$, then $(A-B)-C=(A-C)-B$.

To prove that $(S; t)$ is an abelian group, we define *addition*[37]:

$$A+B=0-((0-A)-B)$$

for $A$ and $B$ in $(S; t)$.

It is obvious that the sum of two elements in $(S; t)$ is a uniquely determined element in $(S; t)$; and the commutativity of addition is a consequence of (B). Next one infers from these rules that

$$\begin{aligned} A + (B - A) &= 0 - ((0 - A) - (B - A)) = 0 - ((0 - (B - A)) - A) \\ &= 0 - (((B - B) - (B - A)) - A) \\ &= 0 - (((B - (B - A)) - B) - A) \\ &= 0 - (((B - (B - A)) - A) - B) \\ &= 0 - (((B - A) - (B - A)) - B) = 0 - (0 - B) = B, \end{aligned}$$

so that $B-A$ is a solution of the equation $A+X=B$. Finally one verifies the associative law of addition as follows

$$\begin{aligned} &A + (B + C) \\ &\quad = 0 - ((0 - A) - (B + C)) = 0 - ((0 - A) - (0 - ((0 - B) - C))) \\ &\quad = 0 - ((0 - (0 - ((0 - B) - C))) - A) = 0 - (((0 - B) - C) - A) \\ &\quad = 0 - (((0 - B) - A) - C) = 0 - (((0 - A) - B) - C) \\ &\quad = 0 - ((0 - (0 - ((0 - A) - B))) - C) = 0 - ((0 - (A + B)) - C) \\ &\quad = (A + B) + C; \end{aligned}$$

and this completes the proof of Lemma III.4.7.

Lemma III.4.8. *If $t<u\leqq g$, and if the maximum order of the subcycles of $u$ does not exceed $m$, then there exists a vector $V$ over $S$ such that $c(V)\leqq u$, though $c(V)$ is not part of $t$.*

**Proof.** Since the parts of $g$ are sums of cycles, there exists a cycle $z$ such that $z\leqq u$, though $z$ is not part of $t$. From our hypothesis it follows that

---

(37) See in this context the following treatments of the postulates of subtraction in abelian groups: M. Ward, these Transactions, vol. 32 (1930), pp. 520–526; D. G. Rabinow, American Journal of Mathematics, vol. 59 (1937), pp. 211–224, 385–392; B. A. Bernstein, these Transactions, vol. 43 (1938), pp. 1–6.

$n(z) \leqq m$. But we proved as a corollary to Theorem III.3.2 the existence of a vector $V$ over $S$ such that $c(V) = z$; and this is just the statement to be proved.

LEMMA III.4.9. *If $z_1, \cdots, z_k$ are subcycles of $g$, and if $V$ is a vector over $S$ such that $c(V) \leqq z_1 + \cdots + z_k$, then there exist vectors $V_i$ over $S$ such that $c(V_i) \leqq z_i$ and $V = V_1 + \cdots + V_k$ (using addition as in Lemma* III.4.7).

For the proof of this theorem we shall need the following lemma.

LEMMA III.4.10. *If $u$ and $v$ are cycles such that $uv = 0$ and such that $u+v$ is primary and splits completely, and if the elements $u$, $v$, $t$ are collinear, then there exists a subcycle $d$ of $t$ such that $u$, $v$, $d$ are collinear.*

**Proof of Lemma** III.4.10. If $t = u+v$, then follows from Lemma I.3.8 the existence of a cycle $d \leqq t$ which is not a subcycle of any proper partial sum of $u+v$ and $u$, $v$, $d$ are clearly collinear.

Thus we assume now that $t < u+v$; and we may assume without loss in generality that $0 < n(u) \leqq n(v) = k$. Since $(u+v)/u$ is a cycle of order $k$, there exists between $u$ and $u+v$ an element $r$ such that $(u+v)/r$ is a cycle of order 1. Since $t+u = u+v$, it follows that $t \leqq r$ does not hold; and since $t$ is a sum of cycles, there exists a subcycle $d$ of $t$ which is not part of $r$. Thus $d+u \leqq r$ does not hold and this implies $d+u = u+v$. Since the orders of the subcycles of $u+v$ do not exceed $k$—by Theorem I.2.1—and since $(u+v)/u$ is a cycle of order $k$, it follows that $n(d) = k$ and $du = 0$. Since $t$ splits as a part of $u+v$, and since $d$ is a subcycle of maximum order of $t$ (and $u+v$), $t$ is the direct sum of $d$ and of a cycle $e$ of order $j$; and $t < u+v = u+d$ implies $j < n(u)$. The cross-cut $dv$ is a cycle of order $i$; and $d+v$ is—as before—the direct sum of $v$ and of a cycle $z$ of order $k-i$. Since $u+v = e+d+v = e+z+v$, and since $n(e) < n(u)$, it follows from Corollary I.3.4 that $k-i = n(z) \not< n(u)$ or $n(u) \leqq n(z)$ and hence $d+v = z+v = u+v$ is a consequence of Corollary I.3.4 so that $u$, $v$, $d$ are collinear.

**Proof of Lemma** III.4.9. Since Lemma III.4.9 is certainly true for $k = 1$, we may assume that it holds true for vectors $W$ such that $c(W) \leqq \sum_{i=1}^{k-1} z_i = s$. Since $z_k$ is a cycle, there exists a uniquely determined subcycle $z$ of $z_k$ such that $s + c(V) = s + z$—note $c(V) \leqq s + z_k$. It follows from Lemma III.1.1 that $c(V)$, $z$ and $s(c(V)+z) = t$ are collinear. From $5 < n$ we infer the existence of an integer $i$ such that $c(i)(z + c(V)) = 0$. The $i$th coordinate $(i, V)$ of the vector $V$ is therefore a well determined cycle. If $r = (c(i)+z)(t + (i, V))$, then it follows from Lemma III.1.2 that the triplets $r$, $c(i)$, $z$ and $r$, $t$, $(i, V)$ are collinear, since the triplets $c(i)$, $(i, V)$, $c(V)$ and $z$, $t$, $c(V)$ are collinear. Since $c(i)z = 0$, it follows from Lemma III.4.10 that there exists a subcycle $d$ of $r$ such that $d$, $c(i)$, $z$ are collinear. By Theorem III.3.2 there exists one and only one vector $B$ such that $c(B) = z$, $(i, B) = d$.

We note $c(B) \leqq z_k$. Since furthermore

$$\begin{aligned} c(V-B) &= (c(V)+c(B))((i,V)+(i,B)) = (c(V)+z)((i,V)+d) \\ &\leqq (c(V)+z)((i,V)+r) = (c(V)+t)((i,V)+t) \\ &= t+(i,V)(c(V)+z) = t, \end{aligned}$$

$V-B$ is a vector such that $c(V-B)\leqq z_1+\cdots+z_{k-1}$; and hence there exist by our induction hypothesis vectors $V_i$ such that $c(V_i)\leqq z_i$ for $i=1,\cdots,k-1$ and such that $V-B=V_1+\cdots+V_{k-1}$ or

$$V=V_1+\cdots+V_{k-1}+B,$$

as was to be shown.

5. **The existence of primary abelian operator groups.** The term *primary abelian operator group* shall signify an abelian group $G$ and a ring $E$ of endomorphisms of $G$ satisfying the following conditions.

(a) Every right-ideal and every left-ideal in $E$ is two-sided.

(b) There exists an ideal $P$ different from $E$ in $E$ such that every ideal different from 0 and $E$ in $E$ is a power of $P$.

(c) $E$ contains every $D(G;E)$-admissible endomorphism of $G$ (where $D(G;E)$ denotes the Dedekind set of all the $E$-admissible subgroups of $G$).

(d) To every element $x$ in $G$ there exists a positive integer $i$ such that $xP^i=0$.

Suppose now that the element $g$ of some partially ordered set has the property.

(D, 6) If $g$ contains one subcycle of order $n$, then $g$ contains at least six independent subcycles of order $n$.

It is an immediate consequence of Theorem II.3.1 that the parts of such an element $g$ form the system of admissible subgroups of essentially at most one primary abelian operator group.

THEOREM III.5.1. *If the element g in a partially ordered set satisfies* (D, 6), *then the following conditions are necessary and sufficient for the parts of g to be the admissible subgroups of some primary abelian operator group.*

(A) *The parts of g form a Dedekind set.*

(B) *Sums of a finite number of subcycles of g split completely and are primary.*

(C) *If $M$ is a nonvacuous set of subcycles of g and contains every cycle $z$ which is contained in the sum of a finite number of cycles in $M$, then there exists one and only one part $s(M)$ of g such that $M$ is exactly the set of subcycles of $s(M)$.*

**Proof.** The necessity of (A) is a well known fact in the theory of abelian operator groups, the necessity of (B) is a consequence of Theorems II.2.1, II.2.4 and I.3.7, and the necessity of (C) is a consequence of the fact that by Theorem II.2.1 every element in a primary abelian operator group generates a cycle.

To prove the sufficiency of (A), (B), (C) we show first that:

(+) *The validity of* (A), (B), (C) (*and* (D, 6)) *implies the existence of a primary ring* $D_0$ *of subgroups of an abelian group* $G$ *and of a projectivity of the set of parts of* $g$ *upon* $D_0$.

*Case* 1. The orders of the subcycles of $g$ are bounded.

Then we denote by $m$ the maximum order of the subcycles of $g$. By (D, 6) there exist 6 independent subcycles of order $m$ of $g$; and hence it follows from (A), (B) and Lemma III.3.1 that there exists a complete 6-$m$-simplex $S$ with vertices and links parts of $g$. If $x \leqq g$, then we denote by $(S; x)$ the set of all those vectors $V$ over $S$ which satisfy $c(V) \leqq x$. It is a consequence of Lemma III.4.7 that $(S; x)$ is an abelian group (with regard to the subtraction and addition introduced in (III.4.4)) whenever $x$ is the sum of a finite number of cycles. From this fact one infers immediately that every $(S; x)$ for $x \leqq g$ is an abelian group and moreover a subgroup of the abelian group $G = (S; g)$. If $x$ and $y$ are different parts of $g$, then it follows from (C) that one of them, say $x$, contains a cycle which the other one does not contain; and thus follows from Lemma III.4.8 the existence of a vector $V$ such that $c(V) \leqq x$ though $c(V) \leqq y$ does not hold, that is, $(S; x) \neq (S; y)$ is a consequence of $x \neq y$. But this fact puts into evidence that mapping the part $x$ of $g$ upon the subgroup $(S; x)$ of $G$ constitutes a projectivity of the Dedekind set of the parts of $g$ upon the system $D_0$ of subgroups $(S; x)$ of the abelian group $G$.—If $x \leqq g$, then denote by $M(x)$ the set of all the subcycles of $x$. It follows from (C) that $x = s(M(x))$; and if $M$ is a set of subcycles of $g$, meeting the requirements of (C), then $M = M(s(M))$. If $J$ is any set of parts of $g$, then let $H$ be the cross-cut of all the sets $M(x)$ for $x$ in $J$. Clearly $s(J)$ is the greatest part of $g$ which is contained in all the $x$ in $J$; and $(S; s(J))$ is the cross-cut of all the $(S; x)$ for $x$ in $J$. Denote furthermore by $K$ the set of all the cycles which are contained in the sum of a finite number of cycles from the set $M(x)$ for $x$ in $J$. This set $K$ meets the requirements of (C) and thus $s(K)$ is a well determined part of $g$. Clearly $(S; s(K))$ contains all the subgroups $(S; x)$ for $x$ in $J$; and it follows from the construction of $K$ and from Lemma III.4.9 that $(S; s(K))$ is exactly the subgroup of $G$ which is generated by all the subgroups $(S; x)$ for $x$ in $J$. Thus $D_0$ has been shown to be a ring of subgroups of $G$. If $V$ is any vector over $S$, then the smallest subgroup of $G$ in $D$ containing $V$ is just $(S; c(V))$; and this is a cycle in the set $D_0$, since the map of $x$ upon $(S; x)$ has been shown to be a projectivity. Thus $D_0$ is a primary ring of subgroups; and this completes the proof of (+) in Case 1.

*Case* 2. The orders of the subcycles of $g$ are not bounded.

Then denote by $M(i)$ the set of all the subcycles of $g$ whose order does not exceed $i$. It is a consequence of Theorem I.2.1 that $M(i)$ meets the requirements of (C); and thus there exists a well determined part $g(i) = s(M(i))$ of $g$ which contains every subcycle of order not exceeding $i$ and none of higher order. Since the orders of the subcycles of $g$ are not bounded, it follows from (D, 6) that $g(i)$ contains at least six independent subcycles of order $i$ (the

maximum order in $g(i)$). Since $g(i) \leqq g$ satisfies (with $g$) the conditions (A), (B), (C) it follows from Case 1 that there exists an abelian group $Q(i)$, a primary ring $D(i)$ of subgroups of $Q(i)$ and a projectivity $q(i)$ of the Dedekind set of the parts of $g(i)$ upon the system $D(i)$ of subgroups. Then $q(i)^{-1}q(i+1)$ is a projectivity of $D(i)$ upon the subgroups of $g(i)^{q(i+1)}$ in $D(i+1)$; and it is a consequence of Theorem II.1.3 that this projectivity is induced by an isomorphism of $Q(i)$ upon the subgroup $g(i)^{q(i+1)}$ of $Q(i+1)$ (in $D(i+1)$). Using these facts one immediately constructs an abelian group $G$, subgroups $G_i$ of $G$, a primary ring $T_i$ of subgroups of $G_i$ and a projectivity $p_i$ of the Dedekind set of the parts of $g(i)$ upon the ring $T_i$ of subgroups, meeting the following requirements.

(i) $G_i \leqq G_{i+1}$, $T_i \leqq T_{i+1}$, $p_i$ and $p_{i+1}$ coincide on the parts of $g(i)$.

(ii) Every element in $G$ is contained in some $G_i$.

If $S_i$ is a subgroup of $T_i$, and if $S_i \leqq S_{i+1}$, then there exists one and only one subgroup $S$ of $G$ which contains all the elements in the $S_i$ and no further elements; and the set $D_0$ of all these subgroups of $G$ is clearly a primary ring, the smallest primary ring containing all the $T_i$.—If $x$ is any part of $g$, then it follows from (C) that $x$ is completely determined by the products $xg(i)$; and if $x(i) \leqq g(i)$, $x(i) \leqq x(i+1)$, then the existence of a smallest part $x$ of $g$, containing all the $x(i)$, is readily inferred from (C). Thus it follows easily that there exists one and only one projectivity $p$ of the set of parts of $g$ upon $D_0$ which coincides with $p_i$ on the parts of $g(i)$; and this completes the proof of (+) in Case 2.

If $G$ is an abelian group and $D_0$ a primary ring of subgroups of $G$ such that there exists a projectivity of the parts of $g$ upon $D_0$, then it follows from (D, 6) and Theorem II.1.2 that $D_0$ is the ring $D(G; E)$ of all the $E$-admissible subgroups of $G$ where $E$ is the ring of all the $D_0$-admissible endomorphisms of $G$; and it is a consequence of Theorems II.2.2 and II.2.3 together with the statements (ii), (v) in §II.2, that the right-ideals in $E$ are two-sided; and that every two-sided ideal different from 0 in $E$ is a power of the uniquely determined prime ideal $P$ in $E$. It is a consequence of (B), (C), Theorems I.3.7 and II.2.4 that every left-ideal in $E$ is two-sided; and thus we have shown that $D_0$ is the set of all the admissible subgroups of the primary abelian operator group $G$ over $E$; and this completes the proof.

Remark. The condition (D, 6) entering into our formulation of the Theorem III.5.1 is patently not necessary for the existence of the primary abelian operator group $G$ over $E$. If we substitute for (D, 6) the condition

(D) *$g$ is contained in an element which satisfies* (A), (B), (C), (D, 6),

then it is readily verified that (D) is necessary and sufficient for the existence of the primary abelian operator group $G$ over $E$.

It is finally an obvious consequence of Theorem II.6.3 that one has to add the conditions (i), (ii) of Theorem II.6.3 to these conditions (A), (B),

(C), (D, 6) (or (D)), in order to assure that the Dedekind set of the parts of $g$ is essentially the same as the set of *all* the subgroups of a suitable primary abelian group.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

# SUFFICIENT CONDITIONS FOR A WEAK RELATIVE MINIMUM IN THE PROBLEM OF BOLZA

BY

E. J. McSHANE

1. **Introduction.** The past decade has brought forth great advances in the theory of the Bolza problem in the calculus of variations and in the theory of the problems (Lagrange, Mayer, and so on) subsumed under it. Ten years ago the necessary conditions of Weierstrass, Clebsch and Jacobi (or Mayer) were established only for minimizing curves normal on every subarc, while the sufficiency theorems needed even more drastic normality assumptions. Now the sufficiency theorems are established under the assumption that Lagrange multipliers $\lambda_0 \geqq 0$, $\lambda_1(x), \cdots, \lambda_m(x)$ exist with which the curve $E_{12}$ satisfies the Euler equation, the transversality condition, and the strengthened Weierstrass, Clebsch and Jacobi conditions. The Euler equation, transversality condition, Weierstrass condition and Clebsch condition are proved necessary with no normality assumptions. Yet normality requirements have not been entirely dispensed with. I have shown(1) that for minimizing curves with order of abnormality 0 or 1 there are multipliers with which all the standard necessary conditions are satisfied. But an example shows that minimizing curves exist, having order of abnormality 2, which do not satisfy all the standard necessary conditions with any multipliers. Thus if the gap is to be closed and the necessary and sufficient conditions brought together for abnormal problems, our only hope is to strengthen the very strong sufficiency theorems of Hestenes, Morse, and Reid.

In a paper to be published in the American Mathematical Monthly, I have considered the problem of minimizing a function $f^0(x) \equiv f^0(x^1, \cdots, x^n)$ subject to conditions

$$(1.1) \qquad f^\beta(x) = 0 \qquad (\beta = 1, \cdots, m).$$

Subject to fairly obvious conditions of definition and differentiability, I have shown that the following condition is necessary in order that $f^0$ have a minimum subject to (1.1) at a point $x_0$ satisfying (1.1).

(N) *To each set* $(u^1, \cdots, u^n)$ *satisfying the conditions*(2)

$$(1.2) \qquad f^\beta_{x^i}(x_0)u^i = 0 \qquad (\beta = 1, \cdots, m)$$

Presented to the Society, May 2, 1941; received by the editors April 4, 1941, and, in revised form, October 11, 1941.

(1) *On the second variation in certain abnormal problems of the calculus of variations*, American Journal of Mathematics, vol. 63 (1941), pp. 516–530.

(2) We use the summation convention, summing on all repeated indices.

*there correspond multipliers* $l_0 \geqq 0, l_1, \cdots, l_m$ *not all zero such that*

$$l_\alpha f^\alpha_{x^i}(x_0) = 0 \tag{1.3}$$

*and*

$$l^\alpha f_{x^i x^j}(x_0) u^i u^j \geqq 0. \tag{1.4}$$

Correspondingly, we show that the following condition is sufficient for $f^0(x)$ to have a proper minimum subject to (1.1) at a point $x_0$ satisfying (1.1).

(S) *To each set of numbers* $(u^1, \cdots, u^n)$ *not all zero satisfying* (1.2) *there corresponds a set of multipliers* $l_0 \geqq 0, l_1, \cdots, l_m$ *such that* (1.3) *holds and the left member of* (1.4) *is positive.*

Here we have no normality assumptions whatever, and still the gap between conditions (N) and (S) is no greater than that between the necessary condition and the sufficient conditions for a minimum of a function $f(x)$ of a single real variable, without side conditions. The distinctive feature of conditions (N) and (S) is the dependence of the multipliers $l_0, \cdots, l_m$ on the solutions $u^1, \cdots, u^n$ of equations (1.2).

Let us now consider the Bolza problem of minimizing a functional

$$g(x_1, y(x_1), x_2, y(x_2)) + \int_{x_1}^{x_2} f(x, y, y')dx \tag{1.5}$$

in the class of functions

$$y^i = y^i(x) \qquad (i = 1, \cdots, n; x_1 \leqq x \leqq x_2) \tag{1.6}$$

satisfying certain differential equations

$$\phi^\alpha(x, y, y') = 0 \qquad (\alpha = 1, \cdots, m < n; x_1 \leqq x \leqq x_2) \tag{1.7}$$

and certain end conditions

$$\psi^\mu(x_1, y(x_1), x_2, y(x_2)) = 0 \qquad (\mu = 1, \cdots, p \leqq 2n + 2). \tag{1.8}$$

Under the usual hypotheses on the functions, we are led by the theorems of the preceding paragraph to the following conjectures.

Conjecture (N). *If a curve*

$$E_{12}: \quad y^i = y^i(x), \qquad x_1 \leqq x \leqq x_2, \tag{1.9}$$

*minimizes the functional* (1.5) *in the class of curves satisfying* (1.7) *and* (1.8) *then for each set of functions* $\eta^i(x)$ *and numbers* $\xi_1, \xi_2$ *which satisfy the equations of variation of* (1.7) *and* (1.8) *there are multipliers* $\lambda^0 \geqq 0, \lambda^\alpha(x)$ *not all zero such that for the function*

$$F(x, y, y', \lambda) \equiv \lambda^0 f(x, y, y') + \lambda^\alpha(x)\phi^\alpha(x, y, y') \tag{1.10}$$

*the Euler equations, transversality condition, Weierstrass condition and Clebsch condition are satisfied, and the second variation formed in the usual way from F and* $\eta$ *and the end functions is non-negative.*

CONJECTURE (S). *In order that a smooth curve* (1.9) *shall give a strong proper relative minimum to the functional* (1.5) *in the class of curves satisfying* (1.7) *and* (1.8), *it is sufficient that the following condition be satisfied. To each nonidentically zero set* $[\eta^i(x), \xi_1, \xi_2]$ *satisfying the equations of variation of* (1.7) *and* (1.8) *there shall correspond multipliers* $\lambda^0 \geqq 0$, $\lambda^\alpha(x)$ *with which the Euler equation, transversality condition and strengthened Weierstrass and Clebsch conditions hold, and the second variation is positive.*

These conjectures are now being investigated, the first by Miss Mary Jane Cox and the second by Mr. Franklin G. Myers(3). The purpose of the present paper is to establish the analogue of Conjecture (S) for weak relative minima. The proof is made by expansion methods, not as a matter of choice but rather as a matter of necessity. The field theory hardly seems applicable. We cannot even find a conjugate set of accessory extremals; worse, we cannot even set up an accessory problem, because of the dependence of the multipliers on the $\eta^i(x)$.

We shall obtain a sufficiency theorem for the parametric form of the problem, and from this we shall deduce a theorem for the non-parametric problem.

2. **Statement of the problem.** We shall study the Bolza problem in parametric form. On an open set $R_1$ of points $(y, r) = (y^0, \cdots, y^n, r^0, \cdots, r^n)$ in $(2n+2)$-dimensional space we are given functions

$$f(y, y'), \qquad \phi^\beta(y, y') \qquad (\beta = 1, \cdots, m < n)$$

of class $C^2$. We assume that if $(y, r)$ is in $R_1$ so is $(y, kr)$ for all $k > 0$, and that $f$ and the $\phi^\beta$ are positively homogeneous of degree 1 in $r$. Also, we are given functions

$$\theta(\alpha), \qquad T^{is}(\alpha) \qquad (i = 0, 1, \cdots, n;\ s = 1, 2)$$

defined and of class $C^2$ on an open set $R_2$ in an $r$-dimensional space of points $(\alpha^1, \cdots, \alpha^r)$.

If $C$ is a rectifiable curve, having a representation

$$C:\quad y^i = y^i(t) \qquad (t_1 \leqq t \leqq t_2;\ i = 0, 1, \cdots, n) \tag{2.1}$$

with absolutely continuous functions $y^i(t)$, and $\alpha$ is a point in $R_2$, we say that the set $(C, \alpha)$ is admissible, or that $C$ is admissible with parameters $\alpha$, if for almost all $t$ the point $(y, \dot{y})$ lies in $R_1$ and satisfies the equations

(3) Added in proof, June 1942: Miss Cox has established Conjecture N; Dr. Myers has shown that the hypotheses of Conjecture S yield a semi-strong minimum, and for certain integrands guarantee a strong minimum.

$$\phi^\beta(y(t), \dot{y}(t)) = 0 \qquad (\beta = 1, \cdots, m), \tag{2.2}$$

and the end conditions

$$y^i(t_s) = T^{is}(\alpha) \qquad (i = 0, 1, \cdots, n; s = 1, 2) \tag{2.3}$$

are satisfied.

The problem of Bolza is the problem of minimizing the functional

$$J(C, \alpha) = \theta(\alpha) + \int_{t_1}^{t_2} f(y(t), y'(t))dt \tag{2.4}$$

on the class of all admissible sets $(C, \alpha)$.

Following Carathéodory, we shall denote by $\dot{y}(t)$ the vector $(y^{0\prime}(t), \cdots, y^{n\prime}(t))$ if this vector is defined and finite, and the vector $(0, \cdots, 0)$ otherwise. We shall denote the length of a vector by enclosing the vector between vertical bars; thus

$$| y_1 - y_2 | = \left\{ \sum_{i=0}^{n} (y_1^i - y_2^i)^2 \right\}^{1/2},$$

$$| \alpha | = \left\{ \sum_{h=1}^{r} (\alpha^h)^2 \right\}^{1/2}.$$

The concept of weak relative minimum has been given several equivalent formulations. For simplicity of notation, let us suppose that $C_0$ is a curve of class $C^1$, represented by equations

$$y^i = y_0^i(t) \qquad (t_1 \leqq t \leqq t_2) \tag{2.5}$$

in which the functions $y^i(t)$ are of class $C^1$ and $|y'| \neq 0$. A curve $C$ is in the *first order $\epsilon$-neighborhood* of $C_0$ if it has a Lipschitzian representation (2.1) such that

$$| y(t) - y_0(t) | < \epsilon \qquad (t_1 \leqq t \leqq t_2) \tag{2.6}$$

and

$$\underset{t_1 \leqq t \leqq t_2}{\text{l.u.b.}} \left| (y(t) - y_0(t))^\cdot \right| / \left| y_0'(t) \right| < \epsilon. \tag{2.7}$$

This neighborhood is easily seen to be independent of the particular representation of $C_0$.

We say that the functional $J(C, \alpha)$ has a *weak relative minimum* at the admissible set $(C_0, \alpha_0)$ if there is a positive number $\epsilon$ such that

$$J(C, \alpha) \geqq J(C_0, \alpha_0)$$

for all admissible sets $(C, \alpha)$ having $C$ in the first order $\epsilon$-neighborhood of $C_0$

and $|\alpha-\alpha_0|<\epsilon$. The minimum is *proper* if equality is excluded from (2.7) except when $(C, \alpha)=(C_0, \alpha_0)$. In this paper we shall set forth conditions which ensure that a set $(C_0, \alpha_0)$ gives a proper weak relative minimum to $J(C, \alpha)$.

3. **Statement of the theorem.** We use a slight modification of the summation convention. The repetition of an index in a term connotes the summation of the values of that term over all values of the repeated index, except that the indices $q$ and $s$ are exempted; we never sum over values of $q$ or of $s$. As usual, for each set of numbers $\lambda^0, \cdots, \lambda^m$ we define

$$F(y, r, \lambda) \equiv \lambda^0 f(y, r) + \lambda^\beta \phi^\beta(y, r). \tag{3.1}$$

Henceforth we suppose that $(C_0, \alpha_0)$ is an admissible set, the curve $C_0$ being represented by (2.5) with functions $y_0^i(t)$ of class $C^1$ which have $|y_0'|>0$. The curve $C_0$ satisfies the *Euler equations* with multipliers $\lambda^0, \lambda^1(t), \cdots, \lambda^m(t)$ if

$$\frac{d}{dt} F_{r^i}(y_0(t), y_0'(t), \lambda) = F_{y^i}(y_0(t), y_0'(t), \lambda) \qquad (i = 0, 1, \cdots, n;\ t_1 \leqq t \leqq t_2). \tag{3.2}$$

The set $(C_0, \alpha_0)$ satisfies the *transversality condition* with multipliers $\lambda^0, \lambda^1(t), \cdots, \lambda^m(t)$ if

$$\begin{aligned} \lambda^0 \theta_h(\alpha) + F_{r^i}(y_0(t_2), y_0'(t_2), \lambda(t_2)) T_h^{i2}(\alpha_0) \\ - F_{r^i}(y_0(t_1), y_0'(t_1), \lambda(t_1)) T_h^{i1}(\alpha_0) = 0 \\ (h = 1, \cdots, r), \end{aligned} \tag{3.3}$$

the subscript $h$ denoting partial differentiation with respect to $\alpha^h$.

For the curve $C_0$ it is well known that the quadratic form

$$F_{r^i r^j}(y_0(t), y_0'(t), \lambda) v^i v^j \tag{3.4}$$

vanishes whenever the vector $v$ is linearly dependent on $y_0'(t)$; that is, whenever there is a number $k$ such that

$$v^i = k y_0^{i\prime}(t) \qquad (i = 0, 1, \cdots, n).$$

The curve $C_0$ is said to satisfy the *strengthened Clebsch condition* if for each $t$ in the interval (3.4) and every vector $v$ linearly independent of $y^{i\prime}(t)$ and satisfying the equations

$$\phi^\beta_{r^i}(y_0(t), y_0'(t)) v^i = 0 \qquad (\beta = 1, \cdots, m) \tag{3.5}$$

the quadratic form (3.4) is positive.

Our definition of admissible variations is somewhat more inclusive than the usual definition. We define an *admissible variation set* to be a set $[\eta(t), u] = [\eta^0(t), \cdots, \eta^n(t), u^1, \cdots, u^h]$ in which the functions $\eta^i(t)$ are ab-

solutely continuous, have derivatives integrable together with their squares, and satisfy the equations of variations of (2.2), that is, the equations

$$\Phi^{\beta}(t, \eta, \dot\eta) \equiv \phi^{\beta}_{y^i}(y_0(t), y_0'(t))\eta^i(t) + \phi^{\beta}_{r^i}(y_0(t), y_0'(t))\dot\eta^i(t) = 0 \qquad (\beta = 1, \cdots, m;\ t_1 \leqq t \leqq t_2), \tag{3.6}$$

for amost all values of $t$ in the interval $[t_1, t_2]$, and the numbers $u^h$ satisfy the equations of variation of (2.3), which are

$$\eta^i(t_s) - T_h^{is}(\alpha_0)u^h = 0 \qquad (s = 1, 2;\ i = 0, 1, \cdots, n). \tag{3.7}$$

If $[\eta, u]$ is an admissible variation set, and $\lambda^0, \cdots, \lambda^m(t)$ are multipliers, we define the second variation due to $[\eta, u]$ by the equation

$$J_2[\eta, u, \lambda] \equiv b_{hk}u^hu^k + \int_{t_1}^{t_2} 2\omega(t, \eta, \dot\eta)dt, \tag{3.8}$$

where

$$\begin{aligned} b_{hk} = \theta_{hk}(\alpha_0) &+ F_{r^i}(y_0(t_2), y_0'(t_2), \lambda(t_2))T^{i2}_{hk}(\alpha_0) \\ &- F_{r^i}(y_0(t_1), y_0'(t_1), \lambda(t_1))T^{i1}_{hk}(\alpha_0) \end{aligned} \tag{3.9}$$

and

$$2\omega(t, \eta, \rho) = F_{y^iy^j}\eta^i\eta^j + 2F_{y^ir^j}\eta^i\rho^j + F_{r^ir^j}\rho^i\rho^j, \tag{3.10}$$

the arguments of the functions in the right member being $(y_0(t), y_0'(t), \lambda(t))$.

An admissible variation set $[\eta, u]$ will be called *essentially null* if

$$u^h = 0 \qquad (h = 1, \cdots, r) \tag{3.11}$$

and there is a function $\rho(t)$ such that

$$\eta^i(t) = \rho(t)y_0^{i\prime}(t) \qquad (t_1 \leqq t \leqq t_2;\ i = 0, 1, \cdots, n). \tag{3.12}$$

For such sets the second variation is known to have the value 0.

We can now state our principal theorem.

THEOREM I. *Let the following hypotheses be satisfied.*

(1) *The set $[C_0, \alpha_0]$ is admissible and the curve $C_0$ is a simple arc of class*(4) *$C^2$, represented by equation (2.5) with functions $y_0^i(t)$ of class $C^2$.*

(2) *For all $t$ in the interval $[t_1, t_2]$ the matrix*

$$\|\phi^{\beta}_{r^i}(y_0(t), y_0'(t))\|$$

*has rank $m$.*

(3) *To each admissible variation set $[\eta, u]$ which is not essentially null there corresponds a set of continuous multipliers $\lambda^0 \geqq 0, \lambda^1(t), \cdots, \lambda^m(t)$ with which*

(4) If we assume $C$ to be of class $C^1$, in the presence of hypothesis (3) we can show by the Hilbert differentiability theorem that it must be of class $C^2$.

*the Euler equations, transversality condition and strengthened Clebsch condition are satisfied, and with which the inequality*

$$J_2(\eta, u, \lambda) > 0 \tag{3.13}$$

*holds.*

*Then* $(C_0, \alpha_0)$ *gives* $J(C, \alpha)$ *a proper weak relative minimum on the class of admissible sets* $(C, \alpha)$.

4. **A change of parameter.** There is no loss of generality in assuming that (2.3) is the representation of $C_0$ in terms of arc length, so that $t_1 = 0$ and $t_2$ is the length of $C_0$ and

$$|\, y_0'(t) \,| = 1 \qquad (t_1 \leqq t \leqq t_2). \tag{4.1}$$

Our proof will be indirect; we assume the theorem false, and arrive (in §9) at a contradiction.

If the theorem is false, there exists a sequence of admissible sets $[C_q, \alpha_q]$ $(q = 1, 2, 3, \cdots)$ with the following properties. Each set $(C_q, \alpha_q)$ is distinct from $(C_0, \alpha_0)$. The curves $C_q$ have Lipschitzian representations

$$C_q\colon\; y^i = y^i_q(t) \qquad (t_1 \leqq t \leqq t_2) \tag{4.2}$$

such that

$$\lim_{q\to\infty} y^i_q(t) = y^i_0(t) \tag{4.3}$$

uniformly in $[t_1, t_2]$ and

$$\lim_{q\to\infty} (y^i_q(t) - y^i_0(t))^{\cdot} = 0 \tag{4.4}$$

uniformly in $[t_1, t_2]$. The $\alpha_q$ satisfy

$$\lim_{q\to\infty} \alpha^h_q = \alpha^h_0 \qquad (h = 1, \cdots, r). \tag{4.5}$$

And finally

$$J(C_q, \alpha_q) \leqq J(C_0, \alpha_0) \qquad (q = 1, 2, \cdots). \tag{4.6}$$

For convenience in our proofs it is highly desirable to choose a particular type of representation of the curves $C_q$. Specifically, we shall prove the following statement.

(4.7) *There is no loss of generality in assuming that the representations* (4.2) *satisfy the equations*

$$[y^i_q(t) - y^i_0(t)]\, y^{i\prime}_0(t) = A_q t + B_q, \tag{4.8}$$

*where* $A_q$ *and* $B_q$ *are constants.*

The range of definition of the functions $y_0^i(t)$ may easily be extended to an interval $t_1-\epsilon<t<t_2+\epsilon$ in such a way that they remain of class $C^2$ and the curve $y^i=y_0^i(t)$ remains a simple arc. Consider the equations

$$[y^i - y_0^i(t)]y_0^{i\prime}(t) - At - B = 0. \tag{4.9}$$

The equations have the initial solutions

$$y^i = y_0^i(t); \quad t_1 \leqq t \leqq t_2; \quad A = B = 0. \tag{4.10}$$

On a neighborhood of the set (4.10) the left member of (4.9) is of class $C^1$; on the set (4.10) the partial derivative of the left member with respect to $t$ has the value 1. Hence by a known theorem on implicit functions the equation (4.9) has a solution

$$t = t(y, A, B) \tag{4.11}$$

defined and of class $C^1$ for all $(y, A, B)$ in a set

$$|A| < \delta, \quad |B| < \delta, \quad |y - y_0(t)| < \delta \qquad (t_1 \leqq t \leqq t_2; \delta > 0) \tag{4.12}$$

and assuming values in the interval $(t_1-\epsilon, t_2+\epsilon)$.

The constants $A_q$, $B_q$ of (4.8) are determined by giving $t$ the values $t_1$, $t_2$. By (4.1) and (4.3) the left member of (4.8) approaches zero uniformly, hence $A_q$ and $B_q$ both tend to zero as $q\rightarrow\infty$. We may therefore assume

$$|A_q| < \delta, \qquad |B_q| < \delta \tag{4.13}$$

for all $q$. For all but a finite number of values of $q$, which we discard, each point $y_q(\tau)$ $(t_1\leqq\tau\leqq t_2)$ lies within a distance $\delta$ of some point of $C$. Hence the functions

$$t_q(\tau) \equiv t(y_q(\tau), A_q, B_q) \qquad (t_1 \leqq \tau \leqq t_2) \tag{4.14}$$

are defined and satisfy the equation

$$[y_q^i(\tau) - y_0^i(t_q(\tau))]y_0^{i\prime}(t_q(\tau)) = A_q t_q(\tau) + B_q. \tag{4.15}$$

The functions $t_q(\tau)$ are clearly Lipschitzian. The constants $A_q$ and $B_q$ were so chosen that equation (4.8) is satisfied at $t_1$ and $t_2$, whence

$$t_q(t_1) = t_1, \qquad t_q(t_2) = t_2. \tag{4.16}$$

By (4.3), $y_q(\tau)$ tends to $y_0(\tau)$ uniformly in the interval $[t_1, t_2]$, so by the definition (4.14) we have

$$\lim_{q\to\infty} t_q(\tau) = t(y_0(\tau), 0, 0) = \tau \tag{4.17}$$

uniformly for $t_1\leqq\tau\leqq t_2$. Since

$$|y_q(\tau) - y_0(t_q(\tau))| \leqq |y_q(\tau) - y_0(\tau)| + |y_0(\tau) - y_0(t_q(\tau))|,$$

relations (4.3) and (4.17) imply

$$\lim_{q\to\infty} \left| y_q(\tau) - y_0(t_q(\tau)) \right| = 0 \tag{4.18}$$

uniformly on $t_1 \leqq \tau \leqq t_2$.

Let $M_q$ be the set of values of $\tau$ in $t_1 \leqq \tau \leqq t_2$ for which $y_q'(\tau)$ is defined; this set constitutes almost all of the interval $t_1 \leqq \tau \leqq t_2$. On this set the function $t_q(\tau)$ also has a derivative, as we see from (4.14). For all $\tau$ in $M_q$ the inequality

$$\left| y_q'(\tau) - y_0'(t_q(\tau)) \right| \leqq \left| y_q'(\tau) - y_0'(\tau) \right| + \left| y_0'(\tau) - y_0'(t_q(\tau)) \right| \tag{4.19}$$

holds. So if $\gamma$ is an arbitrary positive number, for all large $q$ the inequality

$$\left| y_q'(\tau) - y_0'(t_q(\tau)) \right| < \gamma \tag{4.20}$$

holds on $M_q$, as follows from (4.19), (4.4), (4.17) and the continuity of $y_0'$.

By differentiating both members of (4.15) we find that on $M_q$ the equation

$$[A_q + 1 - \{ y_q^i(\tau) - y_0^i(t_q(\tau)) \} y_0^{i\prime\prime}(t_q(\tau))] t_q'(\tau) = y_q^{i\prime}(\tau) y_0^{i\prime}(t_q(\tau)) \tag{4.21}$$

holds. Let $\eta$ be an arbitrary positive number less than 1. Since $A_q$ tends to zero and (4.18) holds, the quantity in square brackets in (4.21) lies between $(1+\eta/2n)^{-1/2}$ and $(1+\eta/2n)^{1/2}$ for all sufficiently large $q$. Since $|y_0'| = 1$, by (4.20) with proper choice of[5] $\gamma$ the right member of (4.21) lies between the same bounds for all large $q$. Hence for all sufficiently large $q$ we have

$$\left(1 + \frac{\eta}{2n}\right)^{-1} < t_q'(\tau) < 1 + \frac{\eta}{2n} \cdot \tag{4.22}$$

In particular, if we choose $\eta = 1$ we find that for all but a finite number of values of $q$ (which we discard from further consideration) the value of $t_q'(\tau)$ exceeds $2/3$ on $M_q$, which is almost all of $t_1 \leqq \tau \leqq t_2$. Hence $t_q(\tau)$ has a Lipschitzian inverse; we denote it by $\tau_q(t)$. By (4.16) we see that

$$\tau_q(t_1) = t_1, \qquad \tau_q(t_2) = t_2.$$

If $N_q$ is the image of $M_q$ under the mapping $t = t_q(\tau)$, then $N_q$ constitutes almost all of $t_1 \leqq t \leqq t_2$. On it the derivative of $\tau_q(t)$ exists and is the reciprocal of the derivative of $t_q(\tau)$, so by (4.22)

$$\left(1 + \frac{\eta}{2n}\right)^{-1} < \tau_q'(t) < 1 + \frac{\eta}{2n} \tag{4.23}$$

for all $t$ in $N_q$, $q$ sufficiently large.

We now show that the equations

$$y^i = y_q^i(\tau_q(t)) \qquad (t_1 \leqq t \leqq t_2)$$

---

(5) The choice $\gamma = 1 - (1+\eta/2n)^{-1/2}$ will serve.

form the desired representation of $C_q$. By (4.15) the equation (4.8) is satisfied. By (4.18), the new functions satisfy the convergence condition (4.3). Let $\eta$ be an arbitrary number between 0 and 1. In (4.20) we choose $\gamma=\eta/4n$. For all large $q$ inequalities (4.20) and (4.23) hold on $N_q$; and $t$ being arc length on $y_0$, we have

$$\begin{aligned}\left|\frac{d}{dt}\left[y_q(\tau_q(t)) - y_0(t)\right]\right| &= \left| y_q'(\tau_q(t))\tau_q'(t) - y_0'(t)\right| \\ &\leqq \left| y_q'(\tau_q(t)) - y_0'(t)\right| \tau_q'(t) + \left|\tau_q'(t) - 1\right| \\ &< \gamma\left(1+\frac{\eta}{2n}\right)+\frac{\eta}{2n} \\ &< \frac{\eta}{n}\cdot\end{aligned}$$

By integrating we see that each component

$$y_q^i(\tau_q(t)) - y_0^i(t)$$

satisfies a Lipschitz condition of constant $\eta/n$, so the derivative has absolute value at most $\eta/n$ where it is defined. Thus

$$\left|\left\{y_q(\tau_q(t)) - y_0(t)\right\}^{\cdot}\right| \leqq \eta \tag{4.24}$$

whenever all the derivatives are defined. Elsewhere inequality (4.24) is trivial, the left member being zero. That is, the left member of (4.24) tends uniformly to zero as $q\to\infty$, completing the proof.

5. **A convergence lemma.** For each $q$ we define a non-negative number $k_q$ by the equation

$$k_q^2 = \left|\alpha_q - \alpha_0\right|^2 + \left(\max\left|y_q(t) - y_0(t)\right|\right)^2 + \int_{t_1}^{t_2}\left|\dot{y}_q(t) - y_0'(t)\right|^2 dt. \tag{5.1}$$

These numbers are actually positive; otherwise $(C_q, \alpha_q)$ would be identical with $(C_0, \alpha_0)$, contrary to hypothesis. Next we define

$$u_q^h = (\alpha_q^h - \alpha_0^h)/k_q \quad (h = 1, \cdots, r; q = 1, 2, \cdots), \tag{5.2}$$

$$\eta_q^i(t) = (y_q^i(t) - y_0^i(t))/k_q \tag{5.3}$$

$$(i = 0, 1, \cdots, n; q = 1, 2, \cdots; t_1 \leqq t \leqq t_2).$$

From the preceding equations we have at once for each $q$ the equation

$$\left|u_q\right|^2 + \left(\max\left|\eta_q(t)\right|\right)^2 + \int_{t_1}^{t_2}\left|\dot{\eta}_q(t)\right|^2 dt = 1, \tag{5.4}$$

so that each summand on the left is at most 1. By the Bolzano-Weierstrass

theorem we can select a subsequence of the $u_q$ which converges to a limit $u_0$. There is no loss of generality in supposing that $\{u_q\}$ is already such a sequence, so that

$$\lim_{q\to\infty} \overset{h}{u}_q = \overset{h}{u}_0 \qquad (h = 1, \cdots, r). \tag{5.5}$$

Let $(\alpha_1, \beta_1), \cdots, (\alpha_l, \beta_l)$ be a set of nonoverlapping subintervals of $[t_1, t_2]$, and let $E$ be the set consisting of the sum of these intervals. Then by the inequality of Schwarz

$$\begin{aligned}\sum_{j=1}^{l} | \eta_q(\beta_j) - \eta_q(\alpha_j) | &= \sum_{1}^{l} \left| \int_{\alpha_j}^{\beta_j} \dot{\eta}_q dt \right| \\ &\leqq \int_E | \dot{\eta}_q | \, dt \leqq \left[ \int_E | \dot{\eta}_q |^2 dt \right]^{1/2} \cdot \left[ \int_E 1 dt \right]^{1/2} \\ &\leqq 1 \cdot [mE]^{1/2} = \left[ \sum_{1}^{l} (\beta_i - \alpha_i) \right]^{1/2}.\end{aligned} \tag{5.6}$$

In particular, letting $l=1$ we see that the $\eta_q$ are equi-continuous. Since they have the uniform bound 1 by (5.4), we know by Ascoli's theorem that the sequence $\{\eta_q\}$ contains a subsequence converging uniformly to a limit function $\eta_0(t)$. We may suppose that $\{\eta_q\}$ is already such a subsequence, so that

$$\lim_{q\to\infty} \overset{i}{\eta}_q(t) = \overset{i}{\eta}_0(t) \qquad (i = 0, 1, \cdots, n) \tag{5.7}$$

uniformly on the interval $[t_1, t_2]$.

If we let $q$ tend to $\infty$ in (5.6) we obtain

$$\sum_{j=1}^{l} | \eta_0(\beta_j) - \eta_0(\alpha_j) | \leqq \left[ \sum_{j=1}^{l} (\beta_j - \alpha_j) \right]^{1/2},$$

so that the functions $\overset{i}{\eta}_0(t)$ are absolutely continuous. We wish now to show that their derivatives have integrable squares.

LEMMA 1([6]). *Under the hypotheses on the $\eta_q$, the squares of the derivatives of the $\overset{i}{\eta}_0(t)$ are summable, and*

$$\int_{t_1}^{t_2} | \dot{\eta}_0 |^2 dt \leqq \liminf_{q\to\infty} \int_{t_1}^{t_2} | \dot{\eta}_q |^2 dt \leqq 1. \tag{5.8}$$

([6]) This is in fact a corollary of almost any theorem on semi-continuity of integrals in non-parametric form. Lemma 3 is also a consequence of known theorems. But it seems preferable to give the fairly simple proofs of these lemmas rather than refer the reader to some exposition containing complications not essential for our present needs.

For each positive integer $k$ we subdivide the interval $[t_1, t_2]$ into $2^k$ equal subintervals by points

$$t_1 = \tau_1 < \tau_2 < \cdots < \tau_{2^k+1} = t_2, \tag{5.9}$$

and define piecewise linear functions $p_k^i(t)$ which coincide with $\eta^i(t)$ at each $\tau_l$ $(l=1, \cdots, 2^k+1)$ and are linear between. Since the derivatives of these functions are constant on each subinterval we find

$$\int_{t_1}^{t_2} |\dot{p}_k|^2 dt = \sum_{l=1}^{2^k} |\eta_0(\tau_{l+1}) - \eta_0(\tau_l)|^2/(\tau_{l+1} - \tau_l). \tag{5.10}$$

By Schwarz' inequality,

$$|\eta_q(\tau_{l+1}) - \eta_q(\tau_l)|^2 \leqq \left[\int_{\tau_l}^{\tau_{l+1}} |\dot{\eta}_q|\, dt\right]^2 \leqq (\tau_{l+1} - \tau_l) \int_{\tau_l}^{\tau_{l+1}} |\dot{\eta}_q|^2 dt.$$

Hence

$$\liminf_{q\to\infty} \int_{\tau_l}^{\tau_{l+1}} |\dot{\eta}_q|^2 dt \geqq |\eta_0(\tau_{l+1}) - \eta_0(\tau_l)|^2/(\tau_{l+1} - \tau_l). \tag{5.11}$$

This, with (5.10), yields

$$\liminf_{q\to\infty} \int_{t_1}^{t_2} |\dot{\eta}_q|^2 dt \geqq \int_{t_1}^{t_2} |\dot{p}_k|^2 dt. \tag{5.12}$$

The integrand on the right is non-negative, and except on the set of measure zero on which one or more of the functions $\eta_0, p_1, \cdots$ are non-differentiable its limit as $k\to\infty$ is $|\dot{\eta}_0(t)|^2$. By Fatou's lemma,

$$\liminf_{q\to\infty} \int_{t_1}^{t_2} |\dot{\eta}_q|^2 dt \geqq \liminf_{k\to\infty} \int_{t_1}^{t_2} |\dot{p}_k|^2 dt \geqq \int_{t_1}^{t_2} |\dot{\eta}_0|^2 dt. \tag{5.13}$$

This establishes the lemma.

6. **The equations of variation.** In order to show that the $\eta_0(t)$ satisfy the equations of variation (3.6) it is convenient to prove a lemma.

Lemma 2. *If $g(t)$ is summable together with its square on the interval $[t_1, t_2]$, then*

$$\lim_{q\to\infty} \int_{t_1}^{t_2} g(t)(\eta_q^i - \eta_0^i)dt = \lim_{q\to\infty} \int_{t_1}^{t_2} g(t)(\dot{\eta}_q^i - \dot{\eta}_0^i)dt = 0 \tag{6.1}$$
$$(i = 0, 1, \cdots, n).$$

The vanishing of the first limit in (6.1) is easily established, for

$$\left|\int_{t_1}^{t_2} g(t)(\eta_q^i - \eta_0^i)dt\right| \leqq \max |\eta_q^i - \eta_0^i| \cdot \int_{t_1}^{t_2} |g|\, dt, \tag{6.2}$$

and the right member tends to zero by (5.9). Consider then the other limit. Let $\epsilon$ be an arbitrary positive number. As is well known, it is possible to find a polynomial $p(t)$ such that

$$\int_{t_1}^{t_2} (g(t) - p(t))^2 dt < \epsilon^2/16. \tag{6.3}$$

By (5.4) and (5.8), we have

$$\left(\int_{t_1}^{t_2} (\dot{\eta}_m^i)^2 dt\right)^{1/2} \leqq 1 \qquad (m = 0, 1, 2, \cdots).$$

So by Minkowski's inequality

$$\left(\int_{t_1}^{t_2} (\dot{\eta}_q^i - \dot{\eta}_0^i)^2 dt\right)^{1/2} \leqq 2 \qquad (q = 1, 2, \cdots).$$

From this and Schwarz' inequality we obtain

$$\begin{aligned} &\left| \int_{t_1}^{t_2} g \cdot (\dot{\eta}_q^i - \dot{\eta}_0^i) dt - \int_{t_1}^{t_2} p \cdot (\dot{\eta}_q^i - \dot{\eta}_0^i) dt \right| \\ &\qquad \leqq \left\{ \int_{t_1}^{t_2} (g - p)^2 dt \right\}^{1/2} \left\{ \int_{t_1}^{t_2} (\dot{\eta}_q^i - \dot{\eta}_0^i)^2 dt \right\}^{1/2} < \epsilon/2. \end{aligned} \tag{6.4}$$

By integration by parts,

$$\int_{t_1}^{t_2} p(t)(\dot{\eta}_q^i - \dot{\eta}_0^i) dt = p(t)(\eta_q^i - \eta_0^i) \Big|_{t_2}^{t_1} - \int_{t_1}^{t_2} p'(t)(\eta_q^i - \eta_0^i) dt. \tag{6.5}$$

The first term on the right tends to zero by (5.7), and the second tends to zero by the part of the lemma already proved. Therefore for all $q$ greater than a certain $q_\epsilon$ we have

$$\left| \int_{t_1}^{t_2} p(t)(\dot{\eta}_q^i - \dot{\eta}_0^i) dt \right| < \epsilon/2. \tag{6.6}$$

Now (6.4) and (6.6) imply

$$\left| \int_{t_1}^{t_2} g(t)(\dot{\eta}_q^i - \dot{\eta}_0^i) dt \right| < \epsilon$$

if $q > q_\epsilon$, and our lemma is established.

The curves $C_0$ and $C_q$ are admissible; so the equations

$$\phi^\beta(y_q(t), \dot{y}_q(t)) = \phi^\beta(y_0(t), y_0'(t)) = 0 \tag{6.7}$$

hold for almost all $t$ in the interval $[t_1, t_2]$. Hence, recalling (5.3), the equation

$$A_q^{\beta,i}(t)\eta_q^i(t) + B_q^{\beta,i}(t)\dot{\eta}_q^i(t) = 0 \tag{6.8}$$

holds for almost all $t$, the coefficients being defined by equations

$$(6.9)\quad \begin{aligned} A_q^{\beta,i}(t) &= \int_0^1 \phi^\beta_{y^i}(y_0(t) + \tau[y_q(t) - y_0(t)],\ y_0'(t) + \tau[\dot{y}_q(t) - y_0'(t)])d\tau, \\ B_q^{\beta,i}(t) &= \int_0^1 \phi^\beta_{r^i}(y_0 + \tau[y_q - y_0],\ y_0' + \tau[\dot{y}_q - y_0'])d\tau. \end{aligned}$$

From (4.3) and (4.4) we deduce

$$(6.10)\qquad \lim_{q\to\infty} A_q^{\beta,i}(t) = \phi^\beta_{y^i}(y_0(t),\ y_0'(t)),$$

$$(6.11)\qquad \lim_{q\to\infty} B_q^{\beta,i}(t) = \phi^\beta_{r^i}(y_0(t),\ y_0'(t))$$

uniformly on $t_1 \leqq t \leqq t_2$. For each such $t$ we have, by (6.8),

$$(6.12)\quad \begin{aligned} &\int_{t_1}^{t} \{A_q^{\beta,i}(t)\overset{i}{\eta}_q(t) + \phi^\beta_{r^i}(y_0,\ y_0')\dot{\overset{i}{\eta}}_q(t)\}dt \\ &\qquad = \int_{t_1}^{t_2} \{\phi^\beta_{r^i}(y_0,\ y_0') - B_q^{\beta,i}(t)\}\dot{\overset{i}{\eta}}_q(t)dt. \end{aligned}$$

By the Schwarz inequality, with (5.4) and (6.11), the right member of (6.12) approaches 0 as $q\to\infty$. The limit of the left member is readily found with the help of (6.10) and Lemma 2; we obtain

$$(6.13)\qquad \int_{t_1}^{t} \{\phi^\beta_{y^i}(y_0(t),\ y_0'(t))\overset{i}{\eta}_0(t) + \phi^\beta_{r^i}\dot{\overset{i}{\eta}}_0\}dt = 0.$$

(Here and henceforth we indicate the arguments only in the first function in a bracketed expression whenever the remaining terms have the same arguments.) By differentiating both members of (6.13) we find that $\eta_0(t)$ satisfies equations (3.6) for almost all $t$ in the interval $t_1 \leqq t \leqq t_2$.

Since the sets $(C_0, \alpha_0)$ and $(C_q, \alpha_q)$ are admissible, the equations

$$(6.14)\qquad \begin{aligned} \overset{i}{y}_0(t_s) &= T^{is}(\alpha_0) \\ \overset{i}{y}_q(t_s) &= T^{is}(\alpha_q) \end{aligned} \qquad (i = 0, 1, \cdots, n;\ s = 1, 2)$$

are satisfied. With (5.2) and (5.3), this implies

$$(6.15)\qquad k_q\overset{i}{\eta}_q(t_s) = T^{is}(\alpha_0 + k_q u_q) - T^{is}(\alpha_0).$$

So by the theorem of mean value there is an $\bar{\alpha}_q$ on the line segment joining $\alpha_0$ and $\alpha_q$ such that

$$(6.16)\qquad \overset{i}{\eta}_q(t_s) = T_h^{is}(\bar{\alpha}_q)\overset{h}{u}_q.$$

By passage to the limit we find

$$\overset{i}{\eta}_0(t_s) = T_h^{is}(\alpha_0)u_0^h \qquad (i = 0, 1, \cdots, n; s = 1, 2). \tag{6.17}$$

Thus it has been shown that $[\eta_0, u_0]$ is an admissible set, as defined in §3.

7. **A semi-continuity proof.** In the course of our proof we shall have need of a generalization of Lemma 1. Let us suppose that $a_{ij}(t)$, $b_{ij}(t)$, $c_{ij}(t)$ $(i, j=0, 1, \cdots, n)$ are functions defined and continuous on the interval $t_1 \leqq t \leqq t_2$. We define

$$I(\eta) = \int_{t_1}^{t_2} [a_{ij}(t)\eta^i(t)\eta^j(t) + 2b_{ij}\eta^i\dot{\eta}^j + c_{ij}\dot{\eta}^i\dot{\eta}^j]dt. \tag{7.1}$$

Concerning such integrals we prove a sequence of three lemmas, the last of which is the one needed later.

LEMMA 3. *If for each $t$ in the interval $[t_1, t_2]$ the quadratic form $c_{ij}v^iv^j$ is positive definite, then*

$$\liminf_{q\to\infty} I(\eta_q) \geqq I(\eta_0). \tag{7.2}$$

It is evident that there is no loss of generality in assuming

$$a_{ij}(t) = a_{ji}(t), \quad c_{ij}(t) = c_{ji}(t) \qquad (t_1 \leqq t \leqq t_2;\ i, j = 0, 1, \cdots, n). \tag{7.3}$$

We readily compute

$$\begin{aligned} a_{ij}\overset{i}{\eta}_q\overset{j}{\eta}_q + 2b_{ij}\overset{i}{\eta}_q\overset{j}{\dot{\eta}}_q + c_{ij}\overset{i}{\dot{\eta}}_q\overset{j}{\dot{\eta}}_q = {} & a_{ij}\overset{i}{\eta}_0\overset{j}{\eta}_0 + 2b_{ij}\overset{i}{\eta}_0\overset{j}{\dot{\eta}}_0 + c_{ij}\overset{i}{\dot{\eta}}_0\overset{j}{\dot{\eta}}_0 \\ & + 2(a_{ij}\overset{j}{\eta}_0 + b_{ij}\overset{j}{\dot{\eta}}_0)(\overset{i}{\eta}_q - \overset{i}{\eta}_0) \\ & + 2(b_{ji}\overset{j}{\eta}_0 + c_{ij}\overset{j}{\dot{\eta}}_0)(\overset{i}{\dot{\eta}}_q - \overset{i}{\dot{\eta}}_0) \\ & + a_{ij}(\overset{i}{\eta}_q - \overset{i}{\eta}_0)(\overset{j}{\eta}_q - \overset{j}{\eta}_0) \\ & + 2b_{ij}(\overset{i}{\eta}_q - \overset{i}{\eta}_0)(\overset{j}{\dot{\eta}}_q - \overset{j}{\dot{\eta}}_0) \\ & + c_{ij}(\overset{i}{\dot{\eta}}_q - \overset{i}{\dot{\eta}}_0)(\overset{j}{\dot{\eta}}_q - \overset{j}{\dot{\eta}}_0). \end{aligned} \tag{7.4}$$

The last term is non-negative, hence

$$\begin{aligned} I(\eta_q) \geqq I(\eta_0) & + 2\int_{t_1}^{t_2} (a_{ij}\overset{j}{\eta}_0 + b_{ij}\overset{j}{\dot{\eta}}_0)(\overset{i}{\eta}_q - \overset{i}{\eta}_0)dt \\ & + 2\int_{t_1}^{t_2} (b_{ji}\overset{j}{\eta}_0 + c_{ij}\overset{j}{\dot{\eta}}_0)(\overset{i}{\dot{\eta}}_q - \overset{i}{\dot{\eta}}_0)dt \\ & + \int_{t_1}^{t_2} a_{ij}(\overset{i}{\eta}_q - \overset{i}{\eta}_0)(\overset{j}{\eta}_q - \overset{j}{\eta}_0)dt \\ & + 2\int_{t_1}^{t_2} b_{ij}(\overset{i}{\eta}_q - \overset{i}{\eta}_0)(\overset{j}{\dot{\eta}}_q - \overset{j}{\dot{\eta}}_0)dt. \end{aligned} \tag{7.5}$$

The second and third terms in the right member approach zero by Lemma 2. The fourth tends to zero by (5.7). The last term can be shown to approach zero by using Schwarz' inequality and recalling (5.7) and (5.4). This establishes the lemma.

LEMMA 4. *If for each $t$ in the interval $[t_1, t_2]$ the quadratic form*

$$c_{ij}v^iv^j$$

*is positive for all non-null vectors $v$ which satisfy the equations*

$$\phi^{\beta}_{r^i}(y_0(t), y_0'(t))v^i = 0 \qquad (\beta = 1, \cdots, m), \tag{7.6}$$

*then inequality* (7.2) *is satisfied.*

Let $E$ be the set of points $(t, v)$ in $(n+2)$-dimensional space satisfying the conditions $t_1 \leqq t \leqq t_2$, $|v| = 1$. For each positive integer $p$ we define $U_p$ to be the set of all points $(t, v)$ at which

$$c_{ij}(t)v^iv^j + p\sum_{\beta=1}^{m}(\phi^{\beta}_{r^i}(y_0(t), y_0'(t))v^i)^2 > 0. \tag{7.7}$$

Every point of $E$ which satisfies (7.6) is in $U_p$ for all $p$; every point of $E$ at which (7.6) is false is in $U_p$ if $p$ is large enough. By the Borel theorem a finite number of the $U_p$, say those with subscripts $p_i, \cdots, p_k$, cover $E$. Let $N$ be the greatest of these subscripts; then

$$c_{ij}(t)v^iv^j + N\sum_{\beta=1}^{m}(\phi_{r^i}v^i)^2 > 0 \tag{7.8}$$

for $t_1 \leqq t \leqq t_2$ and $|v| = 1$. By homogeneity, (7.8) continues to hold if $t_1 \leqq t \leqq t_2$ and $|v| \neq 0$.

Now define

$$I_N(\eta) = \int_{t_1}^{t_2}\left\{a_{ij}\eta^i\eta^j + 2b_{ij}\eta^i\dot{\eta}^j + c_{ij}\dot{\eta}^i\dot{\eta}^j + N\sum_{\beta=1}^{m}(\phi^{\beta}_{y^i}\eta^i + \phi^{\beta}_{r^i}\dot{\eta}^i)^2\right\}dt. \tag{7.9}$$

The terms quadratic in $\dot{\eta}$ constitute the left member of (7.8), so $I_N$ satisfies the hypotheses of Lemma 3, and

$$\liminf I_N(\eta_q) \geqq I_N(\eta_0). \tag{7.10}$$

Since $\eta_0$ satisfies the equations (3.6), the terms added to $I(\eta_0)$ in (7.9) vanish, so that

$$I(\eta_0) = I_N(\eta_0). \tag{7.11}$$

This is not true of $\eta_q$. But from (6.8), with (7.9) and (7.1), we find

$$I_N(\eta_q) - I(\eta_q) = \int_{t_1}^{t_2} N\{\phi^\beta_{y^i}\phi^\beta_{y^j}\eta^i_q\eta^j_q + 2\phi^\beta_{y^i}\phi^\beta_{r^j}\eta^i_q\dot\eta^j_q + \phi^\beta_{r^i}\phi^\beta_{r^j}\dot\eta^i_q\dot\eta^j_q\}dt$$

$$(7.12) \qquad = N\int_{t_1}^{t_2}\{[\phi^\beta_{y^i}\phi^\beta_{y^j} - A^{\beta,i}_q A^{\beta,j}_q]\eta^i_q\eta^j_q$$

$$+ 2[\phi^\beta_{y^i}\phi^\beta_{r^j} - A^{\beta,i}_q B^{\beta,j}_q]\eta^i_q\dot\eta^j_q + [\phi^\beta_{r^i}\phi^\beta_{r^j} - B^{\beta,i}_q B^{\beta,j}_q]\dot\eta^i_q\dot\eta^j_q\}dt.$$

The quantities in square brackets tend uniformly to zero, by (6.10) and (6.11), and the integrals of the absolute values of their coefficients are bounded; so

$$(7.13) \qquad \lim_{q\to\infty}[I_N(\eta_q) - I(\eta_q)] = 0.$$

From (7.10), (7.11) and (7.13) we obtain the conclusion (7.2) of our lemma.

LEMMA 5. *If for each $t$ in the interval $[t_1, t_2]$ the quadratic form $c_{ij}v^iv^j$ is non-negative whenever the vector $v$ is linearly independent of $y_0'(t)$ and satisfies equations* (7.6), *then inequality* (7.2) *is satisfied.*

Let $\epsilon$ be an arbitrary positive number, and let

$$(7.14) \qquad I_\epsilon(\eta) = \int_{t_1}^{t_2}[a_{ij}\eta^i\eta^j + 2b_{ij}\eta^i\dot\eta^j + c_{ij}\dot\eta^i\dot\eta^j + \epsilon\dot\eta^i\dot\eta^i]dt.$$

The quadratic form $c_{ij}v^iv^j$ is here replaced by

$$c_{ij}v^iv^j + \epsilon v^iv^i.$$

This is positive for all non-null vectors $v$ which satisfy equations (7.6). For the second term is positive, while the first is non-negative, by hypothesis if $v$ is linearly independent of $y_0'$ and by continuity if $v$ is a multiple of $y_0'$. So $I_\epsilon$ satisfies the hypotheses of Lemma 4, and

$$(7.15) \qquad \lim I_\epsilon(\eta_q) \geqq I_\epsilon(\eta_0).$$

From (5.4) we deduce

$$(7.16) \qquad 0 \leqq \int_{t_1}^{t_2}\dot\eta^i_q\dot\eta^i_q dt \leqq 1.$$

Hence

$$(7.17) \qquad I(\eta_0) \leqq I_\epsilon(\eta_0) \leqq \liminf_{q\to\infty} I_\epsilon(\eta_q) \leqq \liminf_{q\to\infty} I(\eta_q) + \epsilon.$$

Since $\epsilon$ is an arbitrary positive number, this implies inequality (7.2), and the proof of the lemma is complete.

8. **First case.** We now distinguish two possible cases.

*Case* I. *Either $u_0 \neq (0, \cdots, 0)$ or else the functions $\eta^i_0(t)$ do not all vanish identically.*

*Case* II. *The numbers* $u_0^h$ *are all* 0 *and the functions* $\eta_0^i(t)$ *all vanish identically.*

In this section we shall discuss the first case.

Under the hypotheses of Case I, the set $[\eta, u]$ is not essentially null. This is obvious from the definition if some $u^h$ is not zero. If all the $u^h$ have the value 0, we have by (4.8) and (5.3)

$$\overset{i}{\eta}_q(t)\overset{i\prime}{y}_0(t) = (A_q/k_2)t + (B_q/k_q). \tag{8.1}$$

The left member converges uniformly as $0 \to \infty$, so its limit $\eta_0^i y_0^{i\prime}$ must be linear in $t$. But since $u_0 = 0$ equations (6.17) show that this linear function vanishes at $t_1$ and at $t_2$, so it is identically zero:

$$\overset{i}{\eta}_0(t)\overset{i\prime}{y}_0(t) \equiv 0 \qquad (t_1 \leqq t \leqq t_2). \tag{8.2}$$

Now $\eta_0^i(t)$ cannot have the form $\rho(t)y_0^{i\prime}(t)$ $(i=0, \cdots, n)$; for then (8.2) would give $\rho \equiv 0$, contrary to the hypothesis that the $\eta_0^i$ do not all vanish identically.

By the hypotheses of Theorem I, there are multipliers $\lambda^0 \geqq 0, \lambda^1(t), \cdots, \lambda^m(t)$ with which the set $(C_0, \alpha_0)$ satisfies the Euler equations, transversality condition and strengthened Clebsch condition and with which $J_2(\eta_0, u_0, \lambda)$ is positive. This last condition will not be used until the last paragraph of this section.

Since $\lambda^0$ is non-negative, inequality (4.6) implies

$$\lambda^0 J(C_q, \alpha_q) - \lambda^0 J(C_0, \alpha_0) \leqq 0. \tag{8.3}$$

Recalling that along an admissible curve we have $F = \lambda^0 f$, this can be written in the form

$$\lambda^0\theta(\alpha_q) - \lambda^0\theta(\alpha_0) + \int_{t_1}^{t_2} \{F(y_q, \dot{y}_q, \lambda) - F(y_0, y_0', \lambda)\} dt \leqq 0. \tag{8.4}$$

By Taylor's theorem with integral form of remainder, this implies (with (5.2) and (5.3))

$$\begin{aligned} &\lambda^0\theta_h(\alpha_0)k_q\overset{h}{u}_q + k_q^2\lambda^0 \int_0^1 (1-\tau)\theta_{hk}(\alpha_0 + \tau k_q u_q)\overset{h}{u}_q\overset{k}{u}_q d\tau \\ &\quad + k_q \int_{t_1}^{t_2} \{F_{y^i}(y(t), y'(t), \lambda(t))\overset{i}{\eta}_q + F_{r^i}\overset{i}{\dot{\eta}}_q\} dt \\ &\quad + k_q^2 \int_{t_1}^{t_2}\int_0^1 (1-\tau)\{F_{y^iy^j}(y + \tau k_q\eta_q, y' + \tau k_q\dot{\eta}_q, \lambda)\overset{i}{\eta}_q\overset{j}{\eta}_q \\ &\quad + 2F_{y^ir^j}\overset{i}{\eta}_q\overset{j}{\dot{\eta}}_q + F_{r^ir^j}\overset{i}{\dot{\eta}}_q\overset{j}{\dot{\eta}}_q\} d\tau dt \leqq 0. \end{aligned} \tag{8.5}$$

To the third term in (8.5) we apply the usual integration by parts. Since

by hypothesis the Euler equations are satisfied, this term has the value

$$k_q F_{r^i}\eta_q^i \Big|_{t_1}^{t_2}. \tag{8.6}$$

But

$$\begin{aligned} k_q\eta_q^i(t_s) &= y_q^i(t_s) - y_0^i(t_s) \\ &= T^{is}(\alpha_q) - T^{is}(\alpha_0) \\ &= k_q T_h^{is}(\alpha_0)u_q^h + k_q^2\int_0^1 (1-\tau)T_{hk}^{is}(\alpha_0+\tau k_q u_q)u_q^h u_q^k d\tau. \end{aligned} \tag{8.7}$$

We substitute this in (8.6), and substitute expression (8.6) for the third term in (8.5). Since by hypothesis the transversality condition is satisfied, inequality (8.5) yields, after division by $k_q^2/2$,

$$\begin{aligned} D_q \equiv 2\int_0^1 (1-\tau)\{&\lambda^0\theta_{hk}(\alpha_0+\tau k_q u_q) \\ &+ F_{r^i}(y_0(t_2), y_0'(t_2), \lambda(t_2))T_{hk}^{i2}(\alpha_0+\tau k_q u_q) \\ &- F_{r^i}(y_0(t_1), y_0'(t_1), \lambda(t_1))T_{hk}^{i1}(\alpha_0+\tau k_q u_q)\}u_q^h u_q^k d\tau \\ &+ 2\int_{t_1}^{t_2}\int_0^1 (1-\tau)\{F_{y^i y^j}(y_0+\tau k_q\eta_q, y_0'+\tau k_q\dot\eta_q, \lambda)\eta_q^i\eta_q^j \\ &+ 2F_{y^i r^j}\eta_q^i\dot\eta_q^j + F_{r^i r^j}\dot\eta_q^i\dot\eta_q^j\}d\tau dt \leqq 0. \end{aligned} \tag{8.8}$$

Therefore it is clear that

$$\limsup_{q\to\infty} D_q \leqq 0. \tag{8.9}$$

Since by (5.3)

$$y_0^i(t) + \tau k_q\eta_0^i(t) = y_0^i(t) + \tau[y_q^i(t) - y_0^i(t)],$$

with a like equation for the derivatives, we see by (4.3) and (4.4) that

$$\lim_{q\to\infty} F_{r^i r^j}(y_0+\tau k_q\eta_q, y_0'+\tau k_q\dot\eta_q, \lambda) = F_{r^i r^j}(y_0, y_0', \lambda) \tag{8.10}$$

uniformly for $0\leqq\tau\leqq 1$ and $t_1\leqq t\leqq t_2$. Relations similar to (8.10) hold for the other coefficients in (8.8). Hence if $\epsilon$ is an arbitrary positive number, for all sufficiently large values of $q$ the replacement of $k_q$ by 0 in (8.8) alters each coefficient by less than $\epsilon$. But after this replacement the variable $\tau$ has disappeared from (8.8) except in the factors $(1-\tau)$, whose integral from 0 to 1 has the value 1/2. The result is that the left member of (8.8) takes the form $J_2(\eta_q, u_q, \lambda)$ (cf. (3.8)). Since each coefficient was changed by less than $\epsilon$, we thus find with the help of Schwarz' inequality

$$
(8.11)\qquad
\begin{aligned}
\left| D_q - J_2(\eta_q, u_q, \lambda) \right| &\leqq \epsilon\left( \sum_{h=1}^{r} \overset{h}{u}_q \right)^2 + \epsilon \int_{t_1}^{t_2} \left( \sum_{i=0}^{n} \left\{ \left| \overset{i}{\eta}_q \right| + \left| \dot{\overset{i}{\eta}}_q \right| \right\} \right)^2 dt \\
&\leqq \epsilon\left\{ r \left| u_q \right|^2 + 2(n+1) \int_{t_1}^{t_2} \left( \left| \eta_q \right|^2 + \left| \dot{\eta}_q \right|^2 \right) dt \right\}.
\end{aligned}
$$

The coefficient of $\epsilon$ is bounded, since (5.4) is satisfied, and $\epsilon$ is arbitrary. Hence

$$
(8.12)\qquad \lim_{q\to\infty} \left[ D_q - J_2(\eta_q, u_q, \lambda) \right] = 0.
$$

Relations (8.9) and (8.12) imply

$$
(8.13)\qquad \limsup_{q\to\infty} J_2(\eta_q, u_q, \lambda) \leqq 0.
$$

By (5.5) we have

$$
(8.14)\qquad \lim_{q\to\infty} b_{hk} \overset{h}{u}_q \overset{k}{u}_q = b_{hk} \overset{h}{u}_0 \overset{k}{u}_0.
$$

The integral in (3.8) satisfies the hypotheses of Lemma 3, since the strengthened Clebsch condition holds. Hence from Lemma 3 and equation (8.14) we obtain

$$
(8.15)\qquad \liminf_{q\to\infty} J_2(\eta_q, u_q, \lambda) \geqq J_2(\eta_0, u_0, \lambda).
$$

But now we have reached our desired contradiction. For by the choice of the $\lambda^0$ and $\lambda^\alpha(t)$ the right member of (8.15) is positive, so inequalities (8.13) and (8.15) are incompatible.

9. **Second case.** We still have to dispose of Case II, in which the $u_0$ and $\eta_0(t)$ are all zero. The hypotheses of Theorem I do not mention such variation sets, so we must prove a lemma.

LEMMA 6. *Under the hypotheses of Theorem* I, *there exist admissible variation sets which are not essentially null.*

Bliss(7) has shown that there exist functions $\phi^\gamma(t, r)$ of class $C^2$ ($\gamma = m+1, \cdots, n+1$) such that the determinant

$$
(9.1)\qquad \begin{vmatrix} \phi^{\beta}_{y'^i}(y_0(t), y_0'(t)) \\ \phi^{\gamma}_{y'^i}(t, y_0'(t)) \end{vmatrix}
$$

does not vanish on the interval $t_1 \leqq t \leqq t_2$. In the interval we choose $n+3$ points $\tau_l$ such that

$$
t_1 \leqq \tau_1 < \tau_2 < \cdots < \tau_{n+3} \leqq t_2.
$$

(7) G. A. Bliss, *The problem of Mayer with variable end points*, these Transactions, vol. 19 (1918), pp. 305–314.

For $l=1,\cdots,n+2$ we choose functions $\zeta_l^k(t)$ $(k=1,\cdots,n+1)$ with the following properties. If $k=1,\cdots,m$, then $\zeta_l^k(t)$ is identically zero. Except on $[\tau_l,\tau_{l+1}]$ the other $\zeta_l^k(t)$ also vanish identically. On $[\tau_l,\tau_{l+1}]$ the $\zeta_l^k(t)$ are constants, and the vector

$$(\zeta_l^1(t),\cdots,\zeta_l^{n+1}(t)) \tag{9.2}$$

is linearly independent of

$$(\phi_{r^i}^1 y_0^{i\prime}(\tau_l),\cdots,\phi_{r^i}^{n+1} y_0^{i\prime}(\tau_l)). \tag{9.3}$$

This last condition can be satisfied since by hypothesis $m$ is less than $n$, so that we have the free choice of at least the last two components of the vector (9.2).

By known theorems on differential equations, for $l=1,\cdots,n+2$ the equations

$$\phi_{y^i}^\beta(y_0,y_0')H_l^i+\phi_{r^i}^\beta(y_0,y_0')H_l^{i\prime}=0,\qquad \phi_{r^i}^\gamma(t,y_0')H_l^{i\prime}=\zeta_l^\gamma(t) \tag{9.4}$$
$$(\beta=1,\cdots,m;\ \gamma=m+1,\cdots,n+1)$$

have unique solutions $H_l^i$ vanishing at $t_1$. The $n+1$ homogeneous equations

$$c_lH_l^i(t_2)=0 \qquad (i=0,1,\cdots,n) \tag{9.5}$$

in the $n+2$ unknowns $c_l$ have a non-trivial solution. We define

$$\bar\eta^i(t)=c_lH_l^i(t) \qquad (t_1\leqq t\leqq t_2). \tag{9.6}$$

Then $\bar\eta(t)$ satisfies the equations (3.6), because of (9.4). Since $\bar\eta$ vanishes at $t_1$ and at $t_2$, it satisfies (3.7) with $\bar u=0$. Hence $(\bar\eta,\bar u)$ is an admissible variation set. If it were essentially null, there would be a function $\rho(t)$ such that

$$\bar\eta^i(t)=\rho(t)y_0^{i\prime}(t) \qquad (t_1\leqq t\leqq t_2). \tag{9.7}$$

This $\rho(t)$ is easily seen to be continuous and to have corners only at the points $\tau_l$. If $\lambda$ is the least integer such that $c_\lambda\neq 0$, then $\bar\eta^i(t)$ is identically zero on $[t_1,\tau_\lambda]$. By (9.7) and (9.6)

$$H_\lambda^{i\prime}(\tau_\lambda+)=\bar\eta^{i\prime}(\tau_\lambda+)/c_\lambda=\rho'(\tau_\lambda+)y_0^{i\prime}(\tau_\lambda)/c_\lambda.$$

Since $H_\lambda^i(t_\lambda)$ vanishes, by (9.4) we see that for $l=\lambda$ the vector (9.2) is a multiple of (9.3), contrary to its choice. Lemma 6 is therefore established.

Now by Lemma 6 and hypothesis (3) of Theorem I there are multipliers $\lambda^0\geqq 0,\lambda^1(t),\cdots,\lambda^m(t)$ with which the Euler equations, transversality condition and strengthened Clebsch condition hold. From the last mentioned condition we see that the form

$$F_{r^i r^j}(y_0(t),\ y_0'(t),\ \lambda)v^i v^j \tag{9.8}$$

is positive on the set of unit vectors $v$ which are orthogonal to $y_0'$ and satisfy equations (3.5). This set of vectors is bounded and closed, so on it the form (9.8) has a positive lower bound, which we denote by $2\epsilon$. It follows that on the set of vectors $v$ just described the inequality

$$F_{r^i r^j}v^i v^j - \epsilon v^i v^i > 0 \tag{9.9}$$

holds, since the coefficient of $\epsilon$ has the value 1. By homogeneity, (9.9) continues to hold all non-null vectors $v$ which are orthogonal to $y_0'$ and satisfy equations (3.5).

Let $v$ be any vector which satisfies (3.5) and is linearly independent of $y_0'$. It can be resolved into components

$$v^i = v_0^i + \delta y_0^{i\prime}(t), \tag{9.10}$$

where $v_0$ is orthogonal to $y_0'$. Since $v$ is linearly independent of $y_0'$, the component $v_0$ is not null. The homogeneity of $F$ and $\phi^\beta$ entails the well known consequence

$$y_0^{i\prime}(t)F_{r^i r^j}(y_0,\ y_0',\ \lambda) = 0, \qquad y_0^{i\prime}\phi^\beta_{r^i} = 0. \tag{9.11}$$

Now $y_0'$ and $v$ both satisfy the linear equations (3.5), hence by (9.10) $v_0$ also satisfies those equations. Therefore (9.9) holds with $v_0$ in place of $v$, and with the help of (9.11) we deduce

$$F_{r^i r^j}(y_0,\ y_0',\ \lambda)v^i v^j - \epsilon\left[v^i v^i - (y_0^{i\prime} v^i)^2\right] = F_{r^i r^j}v_0^i v_0^j - \epsilon v_0^i v_0^i > 0. \tag{9.12}$$

Inequality (9.12) shows that the integral

$$I_\epsilon(\eta) = \int_{t_1}^{t_2}\left\{2\omega(t,\ \eta,\ \dot\eta) - \epsilon\left[\dot\eta^i\dot\eta^i - (y_0^{i\prime}\dot\eta^i)^2\right]\right\}dt \tag{9.13}$$

satisfies the hypotheses of Lemma 3, so that

$$\liminf I_\epsilon(\eta_q) \geqq I_\epsilon(\eta_0) = 0. \tag{9.14}$$

Since $u_q$ approaches $u_0=0$ as $q\to\infty$, this and (3.8) together imply

$$\liminf\left\{J_2(\eta_q,\ u_q,\ \lambda) - \epsilon\int_{t_1}^{t_2}\left[\dot\eta_q^i\dot\eta_q^i - (y_0^{i\prime}\dot\eta_q^i)^2\right]dt\right\} \geqq 0. \tag{9.15}$$

By (4.8) and (5.3) the function

$$y_0^{i\prime}(t)\eta_q^i(t)$$

is linear in $t$, and it converges uniformly to zero, since $\eta_0$ is zero. Except on a set of measure zero we have

$$y_0^{i\prime} \dot{\eta}_q^i = (y_0^{i\prime} \eta_q^i)' - y_0^{i\prime\prime} \eta_q^i,$$

and both terms on the right tend uniformly to zero. So the last term in the square bracket in (9.15) can be omitted without affecting the limit. In (5.4) the first and second terms tend to zero, so the third term tends to 1 as $q \to \infty$. Thus (9.15) implies

$$\liminf J_2(\eta_q, u_q, \lambda) - \epsilon \geqq 0. \tag{9.16}$$

On the other hand, the considerations leading, to inequality (8.13) are applicable to Case II as well to Case I, so inequality (8.13) must hold. This contradicts (9.16). Hence in each of the two possible cases we have arrived at a contradiction, and Theorem I is established.

10. **Statement of problem in non-parametric form.** From Theorem I we can deduce its analogue for problems in non-parametric form. We use the formulation due to Morse and Myers[8,9]. The functions

$$f(x, z, p) = f(x, z^1, \cdots, z^n, p^1, \cdots, p^n)$$
$$\phi^\beta(x, z, p) = \phi^\beta(x, z^1, \cdots, z^n, p^1, \cdots, p^n) \qquad (\beta = 1, \cdots, m < n)$$

will be supposed to be defined and of class $C^2$ on an open point set $S_1$ in $(x, z, p)$-space. The functions $T^{is}(\alpha)$ $(i=0, \cdots, n; s=1, 2)$ are defined and of class $C^2$ on an open set $R_2$ in $(\alpha^1, \cdots, \alpha^r)$-space. An admissible set $[z, \alpha]$ is a set of functions $z^i(x)$ absolutely continuous on an interval $[x_1, x_2]$ such that for almost all $x$ in $[x_1, x_2]$ the point $(x, z(x), \dot{z}(x))$ is in $S_1$ and satisfies the equations

$$\phi^\beta(x, z(x), \dot{z}(x)) = 0 \qquad (\beta = 1, \cdots, m), \tag{10.1}$$

together with a set of parameters $\alpha$ in $R_2$ with which the end conditions

$$x_s = T^{0s}(\alpha), \quad z^c(x_s) = T^{cs}(\alpha) \qquad (c = 1, \cdots, n; s = 1, 2) \tag{10.2}$$

are satisfied.

The problem of Bolza in non-parametric form is the problem of minimizing the functional

$$J[z, \alpha] \equiv \theta(\alpha) + \int_{x_1}^{x_2} f(x, z, z')dx \tag{10.3}$$

in the class of admissible sets $[z, \alpha]$.

Let $[z_0, \alpha_0]$ be an admissible set in which the functions

$$z^c = z_0^c(x) \quad (x_{1,0} \leqq x \leqq x_{2,0}; c = 1, 2, \cdots, n) \tag{10.4}$$

---

(8) M. Morse and S. B. Myers, *The problems of Lagrange and Mayer with variable end points*, Proceedings of the American Academy of Arts and Sciences, vol. 66 (1931), pp. 235–253.

(9) M. R. Hestenes, *Sufficient conditions for the problem of Bolza in the calculus of variations*, these Transactions, vol. 36 (1934), pp. 793–818.

are of class $C^1$. The set $[z_0, \alpha_0]$ satisfies the Euler equations with multipliers $\lambda^0, \lambda^1(x), \cdots, \lambda^m(x)$ if the functions

$$F(x, z, p, \lambda) \equiv \lambda^0 f(x, z, p) + \lambda^\beta \phi^\beta(x, z, p) \tag{10.5}$$

satisfy the equations

$$\frac{d}{dx} F_{p^c}(x, z_0, z_0', \lambda) = F_{z^c}(x, z_0, z_0', \lambda) \qquad (c = 1, \cdots, n). \tag{10.6}$$

It satisfies the transversality conditions with multipliers $\lambda^0, \lambda^1(x), \cdots, \lambda^m(x)$ if

$$\lambda^0 \theta_h(\alpha_0) + \left[(F - z_0^{c\prime} F_{p^c}) T_h^{0s} + F_{p^c} T_h^{cs}\right]_1^2 = 0 \qquad (h = 1, \cdots, r). \tag{10.7}$$

As usual, $c$ has the range $1, \cdots, n$. The square-bracketed symbol in (10.7) is to be understood as follows. The functions $F$, and so on are first evaluated at $(x_{s,0}, z_0(x_{s,0}), z_0'(x_{s,0}), \lambda(x_{s,0}))$, $s = 1, 2$. Then the value of the sum inside the square bracket is evaluated for $s = 1$ and for $s = 2$, and the former value subtracted from the latter.

The set $[z_0, \alpha_0]$ satisfies the *strengthened Clebsch condition* with multipliers $\lambda^0, \lambda^1(x), \cdots, \lambda^m(x)$ if for each $x$ in $[x_{1,0}, x_{2,0}]$ the inequality

$$F_{p^c p^d}(x, z_0(x), z_0'(x), \lambda(x)) v^c v^d > 0 \tag{10.8}$$

(summed over $c, d = 1, \cdots, n$) holds for all sets of numbers $(v^1, \cdots, v^n) \neq (0, \cdots, 0)$ satisfying the equations

$$\phi^\beta_{p^c}(x, z_0(x), z_0'(x)) v^c = 0 \qquad (\beta = 1, \cdots, m). \tag{10.9}$$

An *admissible variation set* $[\zeta(x), u]$ is a set consisting of $n$ functions, absolutely continuous and having derivatives whose squares are summable over $[x_{1,0}, x_{2,0}]$ and satisfying the equations

$$\phi^\beta_{z^c}(x, z_0, z_0') \zeta^c(x) + \phi^\beta_{p^c}(x, z_0, z_0') \zeta'^c(x) = 0 \qquad (\beta = 1, \cdots, m) \tag{10.10}$$

for almost all $x$, and a set of numbers $(u^1, \cdots, u^r)$ with which the equations

$$\zeta^c(x_{s,0}) = [T_h^{cs}(\alpha_0) - z_0^{c\prime}(x_{s,0}) T_h^{0s}] u^h \qquad (s = 1, 2;\ h = 1, \cdots, r) \tag{10.11}$$

hold.

If $[\zeta, u]$ is an admissible variation set, we define the second variation due to $[\zeta, u]$ by the equation

$$J_2[\zeta, u, \lambda] = b_{hk} u^h u^k + \int_{x_{1,0}}^{x_{2,0}} 2\omega(x, \zeta, \zeta') dx, \tag{10.12}$$

where

$$b_{hk} = \overset{0}{\lambda}\theta_{hk}(\alpha_0) + \Big[(F_x - z_0^{c\prime} F_{z^c}) T_h^{0s} T_k^{0s} + (F - z_0^{c\prime} F_{y^c}) T_{hk}^{0s} \tag{10.13}$$
$$+ F_{z^c}(T_h^{0s} T_k^{cs} + T_k^{0s} T_h^{cs}) + F_{y^c} T_{hk}^{cs}\Big]_1^2$$

and

$$2\omega(x, \zeta, \pi) = F_{z^c z^d}(x, z_0, z_0', \lambda)\zeta^c\zeta^d + 2F_{z^c y^d}\zeta^c\pi^d + F_{y^c y^d}\pi^c\pi^d. \tag{10.14}$$

The concept of a weak relative minimum will be carried over unchanged from the parametric problem. Let the functions (10.4) be of class $C^1$, and consider another set of absolutely continuous functions

$$z^c = z^c(x), \qquad x_1 \leqq x \leqq x_2. \tag{10.15}$$

We can use these functions to define a curve $C$ in $(n+1)$-space by means of the equation

$$x = t, \qquad z^c = z^c(t) \qquad (x_1 \leqq t \leqq x_2,\ c = 1, \cdots, n), \tag{10.16}$$

and likewise for the set (10.4). For these curves the concept of first order $\epsilon$-neighborhood has already been defined in §2, and so has the concept of weak relative minimum.

For problems in non-parametric form we shall establish the following analogue of Theorem I.

THEOREM II. *Let the following hypotheses be satisfied.*

(1) *The set* $[z_0, \alpha_0]$ *defined by* (10.4) *is admissible, and the functions* $z_0^c(x)$ $(c=1, \cdots, n)$ *are of class* $C^2$.

(2) *For each* $x$ *in the interval* $[x_{1,0}, x_{2,0}]$ *the matrix*

$$\left\| \phi_{y^c}^{\beta}(x, z_0(x), z_0'(x)) \right\| \qquad (\beta = 1, \cdots, m;\ c = 1, \cdots, n) \tag{10.17}$$

*has rank* $m$.

(3) *To each nonidentically vanishing admissible variation set* $[\zeta, u]$ *there corresponds a set of absolutely continuous multipliers* $\lambda^0 \geqq 0, \lambda^1(x), \cdots, \lambda^m(x)$ *with which the set* $[z_0, \alpha_0]$ *satisfies the Euler equations, transversality condition and strengthened Clebsch condition, and with which the inequality*

$$J_2[\zeta, u, \lambda] > 0 \tag{10.18}$$

*is satisfied.*

*Then the set* $[z_0, \alpha_0]$ *gives* $J[z, \alpha]$ *a proper weak relative minimum on the class of admissible sets* $[z, \alpha]$.

In the next two sections we shall show that this theorem is in fact a consequence of Theorem I.

11. **Transformation into parametric form.** We prove Theorem II by replacing the non-parametric problem of §10 by an equivalent parametric problem. The symbols $y^0, y^1, \cdots, y^n$ will be used as alternative names for the

$x, z^1, \cdots, z^n$ axes in $(n+1)$-space. In the $(2n+2)$-dimensional space of points $(y, r) = (y^0, \cdots, y^n, r^0, \cdots, r^n)$ we define $R_1$ to be the set of points $(y, r)$ having $r^0 > 0$ and such that $(y^0, y^1, \cdots, y^n, r^1/r^0, \cdots, r^n/r^0)$ is in $S_1$. On $R_1$ we define functions $g(y, r)$, $\psi^\beta(y, r)$ by the equations

$$(11.1) \qquad \begin{aligned} g(y, r) &= r^0 f(y^0, y^1, \cdots, y^n, r^1/r^0, \cdots, r^n/r^0), \\ \psi^\beta(y, r) &= r^0 \phi^\beta(y^0, y^1, \cdots, y^n, r^1/r^0, \cdots, r^n/r^0). \end{aligned}$$

These have the continuity and homogeneity properties specified in §2.

From (11.1) we deduce

$$(11.2) \qquad g(y^0, \cdots, y^n, 1, r^1, \cdots, r^n) = f(y^0, y^n, r^1, \cdots, r^n),$$

from which by differentiation we obtain similar identities for all existing partial derivatives not involving differentiation with respect to $r^0$; for instance,

$$(11.3) \qquad g_{y^0}(y^0, \cdots, y^n, 1, r^1, \cdots, r^n) = f_x(y^0, \cdots, y^n, r^1, \cdots, r^n).$$

Moreover, identities analogous to (11.2) and its corollaries are also valid for each of the functions $\phi^\beta$.

We now have the data needed for setting up the parametric problem of minimizing the functional

$$(11.4) \qquad I(C, \alpha) \equiv \theta(\alpha) + \int_{t_1}^{t_2} g(y(t), \dot{y}(t))dt$$

in the class of sets $(C, \alpha)$ for which the curves $C: y = y(t)$ have $(y, \dot{y})$ in $R_1$ for almost all $t$, satisfy the differential equations

$$(11.5) \qquad \psi^\beta(y, \dot{y}) = 0 \qquad (\beta = 1, \cdots, m)$$

for almost all $t$, and satisfy the end conditions

$$(11.6) \qquad y^i(t_s) = T^{is}(\alpha) \qquad (i = 0, 1, \cdots, n; s = 1, 2).$$

From (11.1) it is evident that if the set $[z, \alpha]$ with functions $z(t)$ defined by (10.15) is admissible, and $C$ is defined by (10.16), then $(C, \alpha)$ is admissible for the parametric problem, and

$$(11.7) \qquad I(C, \alpha) = J[z, \alpha].$$

(The converse also is true; if $(C, \alpha)$ is admissible for the parametric problem, $C$ can be represented in the form (10.16), and the functions $z^c(x)$ thus obtained are admissible with parameters $\alpha$ for the non-parametric problem. But we do not need this.) Theorem II will therefore be established if we can show that its hypotheses imply that the corresponding parametric problem satisfies the hypotheses of Theorem I.

12. **Verification of the hypotheses of Theorem I.** Hypotheses (1) and (2) of Theorem I follow at once from the corresponding hypotheses of Theorem II.

Let $[\eta(t), u]$ be an admissible variation set for the parametric problem. We define

$$\zeta^i(x) = \eta^i(x) - y_0^{i\prime}(x)\eta^0(x) \qquad (i = 0, \cdots, n;\ x_{1,0} \leqq x \leqq x_{2,0}). \tag{12.1}$$

(The notations $x$ and $t$ for the independent variable are interchangeable, because of (10.16); and for the same reason $z^c(x)$ and $y^c(x)$ are identical, $c=1, 2, \cdots, n$.)

If we recall that

$$y_0^{0\prime}(t) \equiv 1 \tag{12.2}$$

a well known consequence of the homogeneity of the $\psi^\beta$ implies (with (3.6) and the analogues of (11.1))

$$\begin{aligned} \phi^\beta_{zc}(x, z_0, z_0')\zeta^c + \phi^\beta_{pc}(x, z_0, z_0')\dot\zeta^c & \\ &= \psi^\beta_{y^i}(y_0, y_0')(\eta^i - y_0^{i\prime}\eta^0) + \psi^\beta_{r^i}(\dot\eta^i - y_0^{i\prime\prime}\eta^0 - y_0^{i\prime}\dot\eta^0) \\ &= -\eta^0[\psi^\beta_{y^i}y_0^{i\prime} + \psi^\beta_{r^i}y_0^{i\prime\prime}] \\ &= -\eta^0[\psi^\beta(y_0, y_0')]' \\ &= 0. \end{aligned} \tag{12.3}$$

So equations (10.10) hold. Equations (10.11) follow at once from (12.1) and (3.7). Therefore, by hypothesis (3) of Theorem II there exist multipliers $\lambda^0 \geqq 0, \lambda^1(x), \cdots, \lambda^m(x)$ with which the Euler equations, transversality condition and strengthened Clebsch condition hold for the non-parametric problem, and also the inequality (10.18).

Let us define

$$G(y, r, \lambda) = \lambda^0 g(y, r) + \lambda^\beta\psi^\beta(y, r). \tag{12.4}$$

The Euler equations for the non-parametric problem constitute the last $n$ of the $n+1$ Euler equations for the parametric problem. But from the homogeneity of $G$ it can be shown that the equation

$$y^{i\prime}(t)\left\{\frac{d}{dt}G_{r^i}(y, y', \lambda) - G_{y^i}(y, y', \lambda)\right\} = 0 \tag{12.5}$$

holds for any admissible curve $y^i = y^i(t)$ of class $C^2$. Since (12.2) holds and the last $n$ of the factors in braces vanish for $y_0(t)$, the first also vanishes, and all $n+1$ Euler equations are satisfied.

From the homogeneity relation

$$G(y, r, \lambda) = r^i G_{r^i}(y, r, \lambda), \tag{12.6}$$

with (12.2) and (11.1), we find that

$$G_{r^0}(y_0, y_0', \lambda) = F(x, z_0, z', \lambda) - z_0^{c\prime} F_{p^c}(x, z_0, z_0', \lambda). \tag{12.7}$$

This shows that the transversality conditions (10.7) imply the transversality conditions (3.3) for the parametric problem.

Let $(v^0, \cdots, v^n)$ be a vector linearly independent of $y_0'(t)$ and satisfying the equations (3.5) with $\psi^\beta$ in place of $\phi^\beta$. Define

$$w^c = v^c - v^0 y_0^{c\prime}(t) \qquad (c = 1, \cdots, n). \tag{12.8}$$

By the analogue of (12.6) and (12.2),

$$\begin{aligned}\phi^\beta_{p^c}(x, z_0, z_0')w^c &= \psi^\beta_{r^i}(v^i - v^0 y_0^{i\prime}) \\ &= \psi^\beta_{r^i} v^i - v^0 \psi^\beta(y_0, y_0') \\ &= 0.\end{aligned}$$

Also, the $w^c$ are not all zero, since $(v^0, \cdots, v^n)$ is not a multiple of $y_0'(t)$. Since by hypothesis the strengthened Clebsch condition holds,

$$F_{p^c p^d}(x, z_0, z_0', \lambda)w^c w^d > 0. \tag{12.9}$$

A consequence of the homogeneity of $G$ is

$$G_{r^i r^j}(y, r, \lambda)r^j = 0. \tag{12.10}$$

From this and (12.8) we have

$$G_{r^i r^j}(y_0, y_0', \lambda)v^i v^j = F_{p^c p^d}(x, z_0, z_0', \lambda)w^c w^d, \tag{12.11}$$

which is positive by (12.9). Therefore the strengthened Clebsch condition holds for the parametric problem.

Let us define

$$2\omega^*(t, \eta, \rho) \equiv G_{y^i y^j}(y_0, y_0', \lambda)\eta^i\eta^j + 2G_{y^i r^j}\eta^i\rho^j + G_{r^i r^j}\rho^i\rho^j. \tag{12.12}$$

This is a symmetric quadratic form in $(\eta, \rho)$, whence

$$2\overset{*}{\omega}(t, \eta, \rho) = \eta^i \overset{*}{\omega}_{\eta^i}(t, \eta, \rho) + \rho^i \overset{*}{\omega}_{\rho^i}(t, \eta, \rho) \tag{12.13}$$

and

$$\bar\eta^i \overset{*}{\omega}_{\eta^i}(t, \eta, \rho) + \bar\rho^i \overset{*}{\omega}_{\rho^i}(t, \eta, \rho) = \eta^i \overset{*}{\omega}_{\eta^i}(t, \bar\eta, \bar\rho) + \rho^i \overset{*}{\omega}_{\rho^i}(t, \bar\eta, \bar\rho). \tag{12.14}$$

(These identities can also be established easily by direct computation.) It is well known that for every set of functions $\eta^i(t)$ of class $C^2$ the equations

$$y_0^{i\prime}(t)\left\{\frac{d}{dt}\overset{*}{\omega}_{\rho^i}(t, \eta, \eta') - \overset{*}{\omega}_{\eta^i}(t, \eta, \eta')\right\} = 0 \tag{12.15}$$

are satisfied identically. We now prove a lemma.

LEMMA 7. *If the functions $\gamma(t)$, $\eta^0(t), \cdots, \eta^n(t)$ are absolutely continuous and the squares of their derivatives are summable, then*

$$
\begin{aligned}
(12.16)\qquad \int_{t_1}^{t_2} \{ (\gamma y_0^{i\prime})^{\cdot} \omega^*_{p^i}(t, \eta, \dot\eta) + (\gamma y_0^{i\prime}) \omega^*_{\eta^i}(t, \eta, \eta') \} dt \\
= \gamma(t) \eta^i(t) G_{y^i}(y_0(t), y_0'(t), \lambda(t)) \Big|_{t_1}^{t_2}.
\end{aligned}
$$

Let us first suppose that the $\eta^i(t)$ are of class $C^2$. By an integration by parts, with use of (12.15), the left member of (12.16) is reduced to

$$
(12.17)\qquad \gamma y_0^{i\prime} \omega^*_{p^i} \Big|_{t_1}^{t_2}.
$$

The analogue of (12.6) holds for $G_{y^i}$, and with the help of this and (12.10) the expression (12.17) transforms into the right member of (12.16). Hence the lemma holds if the $\eta^i(t)$ are of class $C^2$.

By hypothesis, the derivatives of the $\eta^i$ are of class $L^{(2)}$, so by a known property of the Lebesgue integral there is a sequence of polynomials

$$
p_q^{i\prime}(t) \qquad (i = 0, 1, \cdots, n;\ q = 1, 2, \cdots)
$$

such that

$$
(12.18)\qquad \lim_{q\to\infty} \int_{t_1}^{t_2} | \dot\eta - p_q' |^2 dt = 0.
$$

By the inequality of Schwarz, the functions

$$
p_q^i(t) \equiv \eta^i(t_1) + \int_{t_1}^{t} p_q^{i\prime}(t) dt
$$

converge uniformly to $\eta^i(t)$. The coefficients of the form $\omega^*$ are continuous, so

$$
\begin{aligned}
(12.19)\qquad \lim_{q\to\infty} \int_{t_1}^{t_2} \sum_{i=0}^{n} \{ [\omega^*_{\eta^i}(t, p_q, p_q') - \omega^*_{\eta^i}(t, \eta, \dot\eta)]^2 \\
+ [\omega^*_{p^i}(t, p_q, p_q') - \omega^*_{p^i}(t, \eta, \dot\eta)]^2 \} dt = 0.
\end{aligned}
$$

For each $p_q$ the analogue of (12.16) holds, so by Schwarz' inequality

$$
\begin{aligned}
& \int_{t_1}^{t_2} \{ (\gamma y_0^{i\prime})^{\cdot} \omega^*_{p^i}(t, \eta, \dot\eta) + (\gamma y_0^{i\prime}) \omega^*_{\eta^i}(t, \eta, \dot\eta) \} dt \\
(12.20)\qquad & = \lim_{q\to\infty} \int_{t_1}^{t_2} \{ (\gamma y_0^{i\prime})^{\cdot} \omega^*_{p^i}(t, p_q, p_q') + (\gamma y_0^{i\prime}) \omega^*_{\eta^i}(t, p_q, p_q') \} dt \\
& = \lim_{q\to\infty} \gamma p_q^i G_{y^i}(y_0, y_0', \lambda) \Big|_{t_1}^{t_2} \\
& = \gamma \eta^i G_{y^i}(y_0, y_0', \lambda) \Big|_{t_1}^{t_2}.
\end{aligned}
$$

This establishes the lemma.

By definition (3.8),

$$I_2(\eta, u, \lambda) = \overset{0}{\lambda} \overset{*}{b}_{hk} \overset{h}{u} \overset{k}{u} + \int_{t_1}^{t_2} 2\omega^*(t, \eta, \dot{\eta})dt, \tag{12.21}$$

where

$$\overset{*}{b}_{hk} = \overset{0}{\lambda} \theta_{hk} + \left[G_{rt}(t_s) T_{hk}^{is}(\alpha_0)\right]_1^2. \tag{12.22}$$

From (12.13), (12.21) and (12.1) we deduce

$$\begin{aligned} I_2(\eta, u, \lambda) = \lambda^0 \overset{*}{b}_{hk} \overset{h}{u} \overset{k}{u} &+ \int_{t_1}^{t_2} \left\{ \zeta^i \overset{*}{\omega}_{\eta i}(t, \zeta, \dot{\zeta}) + \dot{\zeta}^i \overset{*}{\omega}_{\rho i}(t, \zeta, \dot{\zeta}) \right\} dt \\ &+ \int_{t_1}^{t_2} \left\{ \zeta^i \overset{*}{\omega}_{\eta i}(t, \overset{0}{\eta} y_0', (\overset{0}{\eta} y_0')') + \dot{\zeta}^i \overset{*}{\omega}_{\rho i}(t, \overset{0}{\eta} y_0', (\overset{0}{\eta} y_0')') \right\} dt \\ &+ \int_{t_1}^{t_2} \left\{ \overset{0}{\eta} y_0^{i\prime} \overset{*}{\omega}_{\eta i}(t, \zeta, \dot{\zeta}) + (\overset{0}{\eta} y_0^{i\prime})' \overset{*}{\omega}_{\rho i}(t, \zeta, \dot{\zeta}) \right\} dt \\ &+ \int_{t_1}^{t_2} \left\{ \overset{0}{\eta} y_0^{i\prime} \overset{*}{\omega}_{\eta i}(t, \overset{0}{\eta} y_0', (\overset{0}{\eta} y_0')') \right. \\ &\qquad \left. + (\overset{0}{\eta} y_0^{i\prime})' \overset{*}{\omega}_{\rho i}(t, \overset{0}{\eta} y_0', (\overset{0}{\eta} y_0')') \right\} dt. \end{aligned} \tag{12.23}$$

The integrand in the first integral in the right member is $2\omega^*(t, \zeta, \dot{\zeta})$, which by (12.12) is the same as $2\omega(t, \zeta, \dot{\zeta})$, since $\zeta^0$ vanishes identically. By Lemma 7, the third integral has the value

$$\overset{0}{\eta} \zeta^i G_{y^i}(y_0, y_0', \lambda) \Big|_{t_1}^{t_2}. \tag{12.24}$$

Since $\zeta^0 = 0$, by (12.4) and (11.1) this is equal to

$$\overset{0}{\eta} \zeta^c F_{z^c}(x, z_0, z_0', \lambda) \Big|_{x_{1,0}}^{x_{2,0}}. \tag{12.25}$$

By the same argument, the fourth integral has the value

$$\overset{0}{\eta} \overset{0}{\eta} y_0^{i\prime} G_{y^i}(y_0, y_0', \lambda) \Big|_{x_{1,0}}^{x_{2,0}}. \tag{12.26}$$

By (12.14), the second integral has the same value (12.25) as the third. We substitute these evaluations in (12.23), and for the end values of $\eta^0$ and $\zeta^c$ we substitute the values given by (3.7) and (10.11). On collecting terms and recalling (12.7), (10.12) and (10.13), we find

$$I_2(\eta, u, \lambda) = J_2[\zeta, u, \lambda], \tag{12.27}$$

which is positive by hypothesis.

We have now verified all the hypotheses of Theorem I for the parametric formulation of our problem, so by that theorem the set $(C_0, \alpha_0)$ gives a proper weak relative minimum to $J(C, \alpha)$ on the class of admissible sets $(C, \alpha)$. This immediately implies that $[z_0, \alpha_0]$ gives $J[z, \alpha]$ a proper weak relative minimum on the class of admissible sets $[z, \alpha]$, and Theorem II is established.

13. **A corollary.** If it were not for our unusually inclusive definition of admissible variation set, Theorem II would at once include Hestenes' sufficiency theorem for weak relative minima(10). For Hestenes assumes the hypotheses of Theorem II, with the additional requirement that the multipliers can be chosen independently of the sets $[\zeta, u]$. However, it requires some proof to show that in this case the assumption that the second variation is positive for all variation sets admissible in the sense of §3 is necessarily satisfied if the second variation is positive whenever $[\zeta, u]$ is an admissible set and the $\zeta^c(x)$ are of class $D^1$. We establish this for the normal case; as Bliss has shown(11), the theorem of Hestenes can be deduced from the sufficiency theorem for the normal problem.

For each real number $a$, let us define

$$(13.1) \qquad Q_a[\zeta, u, \lambda] = \lambda^0 b_{hk} u^h u^k + a u^h u^h + \int_{x_{1,0}}^{x_{2,0}} \{2\omega(x, \zeta, \dot\zeta) + a\zeta^c\zeta^c\} dx.$$

Clearly

$$(13.2) \qquad Q_0[\zeta, u, \lambda] = J_2[\zeta, u, \lambda].$$

As in the discussion of (7.7), we can show that if $a$ is sufficiently large the quadratic form

$$(13.3) \qquad \lambda^0 b_{hk} u^h u^k + a u^h u^h$$

is positive definite and

$$(13.4) \qquad 2\omega(x, \zeta, v) + a\zeta^c\zeta^c$$

is positive for all nonidentically zero sets $(\zeta, v)$ satisfying (10.9). We choose such an $a$; then there is a positive $\epsilon$ such that

$$(13.5) \qquad \lambda^0 b_{hk} u^h u^k + a u^h u^h \geqq \epsilon u^h u^h$$

and

$$(13.6) \qquad 2\omega(x, \zeta, \dot\zeta) + a\zeta^c\zeta^c \geqq \epsilon[\zeta^c\zeta^c + \dot\zeta^c\dot\zeta^c]$$

whenever $\dot\zeta$ satisfies (10.9).

Let $K_1$ be the collection of admissible sets $[\zeta, u]$ such that

---

(10) Hestenes, loc. cit., p. 816.

(11) G. A. Bliss, *Normality and abnormality in the calculus of variations*, these Transactions, vol. 43 (1938), pp. 365–376.

$$u^h u^h + \int_{x_{1,0}}^{x_{2,0}} \zeta^e \zeta^e dx = 1. \tag{13.7}$$

For any $[\zeta, u]$ in $K_1$ and any number $b$ we have

$$Q_a[\zeta, u, \lambda] - Q_b[\zeta, u, \lambda] = a - b. \tag{13.8}$$

Since the forms (13.3) and (13.4) are non-negative, $Q_a$ has a non-negative lower bound $m$ on the class $K_1$. Let $[\zeta_q, u_q]$ be a sequence of sets in $K_1$ for which $Q_a$ tends to its lower bound $m$ on $K_1$. By (13.5) and (13.6),

$$Q_a[\zeta, u, \lambda] \geqq \epsilon\left\{u^h u^h + \int_{x_{1,0}}^{x_{2,0}} (|\zeta|^2 + |\dot{\zeta}|^2) dx\right\}, \tag{13.9}$$

so the value of the expression in braces is bounded on the sequence $[\zeta_q, u_q]$. The boundedness of the integral of $|\dot{\zeta}_q|^2$ implies the equi-continuity of the $\zeta_q$, as in (5.6). The boundedness of $|u_q|$ implies the boundedness of $|\zeta_q(x_{1,0})|$, and this with the equi-continuity of the $\zeta_q$ implies their uniform boundedness. By Ascoli's theorem, we can select a subsequence converging uniformly to a limit $\zeta_0(x)$; we suppose $[\zeta_q, u_q]$ such a sequence. We may also suppose that the $u_q$ converge to a limit $u_0$. As in §5, the $\zeta_0^e$ are absolutely continuous, and the squares of their derivatives are summable, and by Lemma 2

$$Q_a[\zeta_0, u_0, \lambda] \leqq \liminf_{q\to\infty} Q_a[\zeta_q, u_q, \lambda] = m. \tag{13.10}$$

But $[\zeta_0, u_0]$ also belongs to $K_1$, so inequality is impossible in (13.10). That is, $[\zeta_0, u_0]$ minimizes $Q_a$ on the class $K_1$.

By (13.8), $[\zeta_0, u_0]$ also minimizes $Q_{a-m}$ on $K_1$, and

$$Q_{a-m}[\zeta_0, u_0, \lambda] = 0.$$

Since $Q_{a-m}$ is homogeneous of degree 2 in $[\zeta, \dot{\zeta}, u]$ we see that $Q_{a-m}$ is non-negative for all admissible variation sets, and on this class $[\zeta_0, u_0]$ minimizes $Q_{a-m}$.

Now we need only a slight extension of Bliss'(12) proof of the multiplier rule to show that there are absolutely continuous multipliers $\mu^0, \mu^1(x), \cdots, \mu^m(x)$ such that for the function

$$2\Omega(x, \zeta, \pi, \mu) = 2\mu^0\omega(x, \eta, \pi) + \mu^\beta[\phi^\beta_{ze}(x, z_0, z_0')\zeta^e + \phi^\beta_{pe}\pi^e]$$

the du Bois-Reymond relation is satisfied:

$$\Omega_{\pi e}(x, \zeta_0, \dot{\zeta}_0) = a_e + \int_{x_{0,1}}^{x} \Omega_{\zeta e}(x, \zeta_0, \dot{\zeta}_0) dx,$$

---

(12) G. A. Bliss, *The problem of Lagrange in the calculus of variations*, American Journal of Mathematics, vol. 52 (1930), pp. 673–744.

where $a_c$ is a constant ($c=1, \cdots, n$). But if the minimizing curve is normal, so is $[\zeta_0, u_0]$. So we may suppose $\mu_0=1$. It follows that

$$\Omega_{r_c r_d}(x, \zeta_0, \dot{\zeta}_0)v^c v^d = F_{p^c p^d}(x, z_0, z_0')v^c v^d,$$

and this last is positive for all nonzero sets $(v^1, \cdots, v^n)$ satisfying (10.9). From this we deduce without trouble that $\dot{\zeta}_0$ must be continuous, so that $\zeta_0$ is of class $C^1$.

Now by (13.2) and (13.8) we have

$$J_2[\zeta_0, u_0, \lambda] = Q_{a-m}[\zeta_0, u_0, \lambda] - a + m = m - a.$$

But since $\zeta_0$ is of class $C^1$, by hypothesis the left member is positive. Hence $m-a$ is positive, and the minimum of $J_2$ on the class $K_1$ is positive. It follows by homogeneity that $J_2$ is positive for all nonidentically vanishing admissible variation sets $[\eta, u]$, as was to be proved.

Since the proofs in this section are rather long, the following remark may not be amiss. Most of the theory of the calculus of variations, originally developed for functions of class $D^1$, can be extended without difficulty to absolutely continuous functions. It seems reasonable to suppose that in the majority of specific problems in which the inequality $J_2[\zeta, u, \lambda]>0$ can be established for functions $\zeta$ of class $D^1$, it will be possible to use essentially the same proof to establish the positiveness of $J_2$ for all variation sets admissible in the sense of §3.

14. **An example.** We now exhibit an example of a problem to which Theorem II applies, but which is not covered by any previously published theorem. A simpler example could be given, but this one will have the virtue of being non-trivial. It is convenient to use subscripts instead of superscripts for enumeration of variables, and so on, since we need to use exponents.

Interior to the interval $[0, 1]$ we choose three disjoint closed intervals $\delta_1, \delta_2, \delta_3$ all of the same length $\epsilon$ (necessarily less than 1/3). It is easy to find functions $\gamma_1(x), \gamma_2(x)$ of class $C^2$ on $[0, 1]$ such that $\gamma_1$ has the value 1 on $\delta_1$ and the value 0 on $\delta_2$ and $\delta_3$, while $\gamma_2$ has the value 1 on $\delta_2$ and the value 0 on $\delta_1$ and $\delta_3$. We define

$$f(x, z_1, \cdots, z_5, p_1, \cdots, p_5) \equiv p_1^2 + p_2^2 + p_3^2 - 4z_1p_1 - 4z_2p_2. \tag{14.1}$$

The problem is to minimize

$$J[z, \alpha] \equiv \int_0^1 f(x, z, z')dx \tag{14.2}$$

in the class of sets $[z_1, \cdots, z_5, \alpha_1, \alpha_2, \alpha_3]$ with absolutely continuous functions $z_i(x)$ ($i=1, \cdots, 5$) which satisfy the end conditions

$$\begin{aligned} &x_1 = z_i(x_1) = 0 \qquad (i = 1, \cdots, 5), \\ &x_2 = 1, \qquad z_4(x_2) = z_5(x_2) = 0, \\ &z_1(x_2) = \alpha_1, \qquad z_2(x_2) = \alpha_2, \qquad z_3(x_2) = \alpha_3, \end{aligned} \tag{14.3}$$

and the differential equations

$$
(14.4)\qquad \begin{aligned}
\phi_1(x, z, z') &\equiv z_4' + z_1 z_1' - z_2 z_2' + \gamma_1(x) z_3'^2 = 0,\\
\phi_2(x, z, z') &\equiv z_5' + z_1 z_2' + z_2 z_1' + \gamma_2(x) z_3'^2 = 0.
\end{aligned}
$$

By use of Theorem II we shall show that the set

$$
(14.5)\qquad z_i(x) \equiv 0, \qquad \alpha_1 = \alpha_2 = \alpha_3 = 0 \qquad (0 \leqq x \leqq 1)
$$

gives $J[z, \alpha]$ a weak relative minimum on the class of admissible sets $[z, \alpha]$.

As usual, we define $F = \lambda_0 f + \lambda_1(x)\phi_1 + \lambda_2(x)\phi_2$. The Euler equations simplify to the form

$$
\lambda_1' = \lambda_2' = 0,
$$

and therefore are satisfied for all sets of constant multipliers. Henceforth we assume $\lambda_1$ and $\lambda_2$ constant. The transversality condition is identically satisfied. The quadratic form (10.8) is

$$
(14.6)\qquad 2\lambda_0(v_1^2 + v_2^2 + v_3^2),
$$

while equations (10.9) become

$$
v_4 = v_5 = 0.
$$

Hence the strengthened Clebsch condition holds if and only if $\lambda_0$ is positive.

Equations (10.10) take the form

$$
\zeta_4' = 0, \qquad \zeta_5' = 0,
$$

and are therefore satisfied by any set $(\zeta_1, \cdots, \zeta_5)$ with $\zeta_4$ and $\zeta_5$ constant. Equations (10.11) become

$$
(14.7)\qquad \begin{aligned}
&\zeta_i(0) = 0 \qquad\qquad (i = 1, \cdots, 5),\\
&\zeta_1(1) = u_1, \quad \zeta_2(1) = u_2, \quad \zeta_3(1) = u_3, \quad \zeta_4(1) = \zeta_5(1) = 0.
\end{aligned}
$$

Thus an admissible variation set is a set $[\zeta, u]$ in which $\zeta_4$ and $\zeta_5$ vanish identically and the other three $\zeta_i$ vanish at $x = 0$ and are absolutely continuous and have derivatives summable with their squares, and in which the $u_h$ satisfy (14.7).

The coefficients $b_{hk}$ are all zero, so if we observe that certain of the terms in $2\omega$ are perfect differentials and make use of (14.7) we obtain

$$
(14.8)\qquad \begin{aligned}
J_2[\zeta, u] = 2\lambda_0 \int_0^1 (\zeta_1'^2 + \zeta_2'^2 + \zeta_3'^2)dx - 4\lambda_0(u_1^2 + u_2^2) + \lambda_1(u_1^2 - u_2^2)&\\
+ 2\lambda_2 u_1 u_2.&
\end{aligned}
$$

If for every nonidentically null admissible variation set we can choose constants $\lambda_0 > 0$, $\lambda_1$ and $\lambda_2$ for which this is positive, all the hypotheses of Theorem II will be verified.

If the derivatives $\dot{\zeta}_i$ vanish almost everywhere, the $\zeta_i$ are constants. In this case, by (14.7) the $\zeta_i$ vanish identically and the $u_i$ also vanish, so $[\zeta, u]$ is identically null. Therefore if $[\zeta, u]$ is not identically null, the integral in (14.8) is positive. If $u_1 = u_2 = 0$, the choice $\lambda_0 = 1$, $\lambda_1 = \lambda_2 = 0$ serves. Otherwise we could choose $\lambda_0 = u_1^2 + u_2^2$, $\lambda_1 = 4(u_1^2 - u_2^2)$, $\lambda_2 = 8u_1u_2$. Then

$$J_2[\zeta, u] = 2(u_1^2 + u_2^2) \int_{t_1}^{t_2} (\dot{\zeta}_1^2 + \dot{\zeta}_2^2 + \dot{\zeta}_3^2) dx > 0.$$

It is easy to show([13]) that it is not possible to choose any one set of multipliers with which the second variation is positive for every nonidentically zero admissible variation set $[\zeta, u]$.

To show that the problem is not a trivial one, in which the extremal (14.5) is isolated, and also that the problem does not impose any hidden end conditions, let us choose any three numbers $\alpha_1, \alpha_2, \alpha_3$. Let $z_1, z_2$ be any Lipschitzian functions vanishing at $x = 0$ and assuming the respective values $\alpha_1, \alpha_2$ at $x = 1$. We determine three numbers, $a_1, a_2, a_3$ by the conditions

$$\begin{aligned} \epsilon a_1^3 + (1/2)(\alpha_1^2 - \alpha_2^2) &= 0, \\ \epsilon a_2^3 + \alpha_1\alpha_2 &= 0, \\ \epsilon(a_1 + a_2 + a_3) - \alpha_3 &= 0. \end{aligned} \tag{14.9}$$

(The number $\epsilon$ was defined in the second paragraph of this section.) Let $\dot{z}_3$ be the function which has the value $a_i$ on the interval $\delta_i$ $(i = 1, 2, 3)$ and is zero elsewhere, and let

$$z_3(x) = \int_0^1 \dot{z}_3(x)dx.$$

By the last of equations (14.9) we have $z_3(1) = \alpha_3$. The functions $z_4, z_5$ are determined by (14.4), with the initial values 0. If we integrate from 0 to 1 in (14.4), by (14.9) we find $z_4(1) = z_5(1) = 0$. Hence the functions $z_i(x)$ satisfy the conditions (14.3) and (14.4). Furthermore, it is clear that they can be made to lie in an arbitrarily small first order neighborhood of (14.5) by restricting $|\dot{z}_1|$ and $|\dot{z}_2|$ to be uniformly small and restricting $\alpha_1, \alpha_2$ and $\alpha_3$ to lie near zero.

15. **Extension to rectifiable curves.** The proofs of our sufficiency theorems did not depend in any essential way on the continuity of the derivatives $y_0'(t)$ or $z_0'(x)$. Theorem I, for example, can be generalized by letting $C_0$ be a rectifiable curve. This of course requires an investigation of the concept of first order neighborhood in the space of rectifiable curves. Such an investiga-

---

([13]) E. J. McShane, *On the second variation in certain normal problems of the calculus of variations*, American Journal of Mathematics, vol. 63 (1941), §5.

tion has already been made(14). Also the formulation of the strengthened Clebsch condition must be altered. The condition as stated in §3 is equivalent, when the functions $y_0(t)$ are of class $C^1$, to the following.

*There is a positive number* $\epsilon$ *such that the inequality*

$$F_{r^i r^j}(y_0(t), y_0'(t), \lambda)v^iv^j \geqq \epsilon |v|^2 \tag{15.1}$$

*holds for all* $t$ *and all vectors* $v$ *orthogonal to* $y_0'(t)$.

It is this latter form which seems appropriate for extension to the case of rectifiable curves. We say that a rectifiable curve $C_0$: $y^i = y_0^i(t)$, $t_1 \leqq t \leqq t_2$, satisfies the strengthened Clebsch condition if (15.1) holds for almost all $t$ such that $y_0'(t)$ is defined and is different from $(0, \cdots, 0)$ and for all $v$ orthogonal to $y_0'(t)$.

(14) E. J. McShane, *Curve-space topologies associated with variational problems*, Annali della R. Scuola Normale Superiore di Pisa, (2), vol. 9 (1940), pp. 45–60.

University of Virginia,
Charlottesville, Va.

# A NEW CLASS OF SELF-ADJOINT BOUNDARY VALUE PROBLEMS

BY

WILLIAM T. REID

1. **Introduction.** Bliss [1](1) has given a general definition of self-adjointness for a differential system of the form

$$(1.1)\qquad \begin{aligned} y_i' &= \sum_{j=1}^{n} [A_{ij}(x) + \lambda B_{ij}(x)]y_j, \\ s_i[y] &\equiv \sum_{j=1}^{n} [M_{ij}y_j(a) + N_{ij}y_j(b)] = 0, \qquad i = 1, \cdots, n. \end{aligned}$$

In the paper above referred to, he has also discussed in detail a special class of self-adjoint problems termed definitely self-adjoint. In a subsequent paper [2], Bliss has given a modification of the definition of definite self-adjointness which is weaker than that previously considered, and has shown that most of the properties deduced in [1] are still valid for systems which are definitely self-adjoint according to the new definition.

If (1.1) is self-adjoint under the nonsingular real transformation $z_i = T_{ij}(x)y_j$(2), in both the original and modified definition of definitely self-adjoint problems Bliss has imposed the definiteness property of the system specifically on the matrix $S(x) \equiv \|S_{ij}(x)\| \equiv \|T_{ki}(x)B_{kj}(x)\|$. Now if $y_i(x)$, $(i=1, \cdots, n)$, is a solution of the differential equations of (1.1) for a value $\lambda$ it follows immediately that

$$(1.2)\qquad \int_a^b y_i T_{ji}[y_j' - A_{jk}y_k]dx = \lambda \int_a^b y_i S_{ij} y_j dx.$$

It may be readily verified that the definiteness property of $S(x)$ assumed by Bliss could equally well have been phrased as a definiteness property of the quadratic functional

$$\int_a^b y_i S_{ij} y_j dx.$$

If $H[y]$ denotes the first member of (1.2), the theory of pencils of quad-

Presented to the Society, September 12, 1940; received by the editors May 14, 1941; corrections received June 17, 1942.

1 Numbers in square brackets refer to the bibliography at the end of this paper.

2 In the introduction, and throughout the paper where matrix notation is not more convenient, the repetition of a subscript in a single term of an expression denotes summation with respect to that subscript over its range of definition.

ratic forms in a finite number of variables suggests that a differential system (1.1) for which $H[y]$ satisfies suitable conditions of definiteness may possess properties analogous to those enjoyed by the class of definitely self-adjoint problems as defined by Bliss [2]. The prime aim of this paper is to show that such is indeed the case. The class of self-adjoint problems herein studied, for which the definiteness property is placed on the functional $H[y]$, is termed $H$-definitely self-adjoint. Moreover, it is to be emphasized that the study of $H$-definitely self-adjoint problems affords new results for systems which are definitely self-adjoint in the sense of Bliss.

The definition of an $H$-definitely self-adjoint system is given in §2, and properties of the functional $H[y]$ are presented in §3. Preliminary results for such a system are obtained in §4; one of the most important results therein contained is that of Theorem 4.3, which states that for an $H$-definitely self-adjoint system the matrix $B(x)$ must be such that its square is identically zero on the interval $ab$. This result, which at first seems startling in aspect, admits certain important consequences for definitely self-adjoint systems. The fundamental properties of an $H$-definitely self-adjoint system, such as the reality of the characteristic values, the equality of the index and multiplicity of a characteristic value, and a type of completeness property of the totality of characteristic solutions for such a system, are contained in §5; §6 is devoted to the discussion of the existence of characteristic values for such a system. Results for definitely self-adjoint systems are given in §7, whereas §8 is concerned with a special definitely self-adjoint problem which is related to a given system (1.1), although the system (1.1) itself may be neither definitely nor $H$-definitely self-adjoint. By the use of the results of the preceding section, extremizing properties of the characteristic values and characteristic solutions of an $H$-definitely self-adjoint system are established in §9. In §10 it is shown that an important instance of the type of boundary value problems associated with the calculus of variations previously studied by the author [9] is $H$-definitely self-adjoint. The connection between the class of problems herein treated and the boundary problems associated with a single linear differential equation of even order which have been studied by Krein [7] and Kamke [6] is indicated briefly in §11. Finally, §12 is devoted to the extension of the notion of $H$-definite self-adjointness to the case of systems whose coefficients are complex-valued.

For simplicity, matrix notation is used almost exclusively in this paper. Square matrices with $n$ rows and columns are denoted by capital italic letters, and the element in the $i$th row and $j$th column is denoted by the letter representing the matrix with the subscript $ij$. Lower case italic letters signify vectors with $n$ components, the $i$th component being signified by a subscript $i$. If $M \equiv \|M_{ij}\|$, $u \equiv [u_i]$, the vectors $[M_{ij}u_j]$ and $[u_jM_{ji}]$ are denoted by $Mu$ and $uM$, respectively. The scalar product $u_jv_j$ of two vectors is written $uv$. If $\alpha$ is a scalar, $\bar{\alpha}$ is its complex conjugate, and for a vector $u$ we write $\bar{u}$

for $[\bar{u}_i]$. For a matrix $M=\|M_{ij}\|$ we use $\tilde{M}$ for the transpose matrix $\|M_{ji}\|$. Finally, if the elements of $M$ are differentiable functions, the matrix of derivatives is denoted by $M'$; similarly, if the components of $u$ are differentiable functions, we write $u'\equiv[u_i']$. The norm of a vector $u$, $[u\bar{u}]^{1/2}$, is written norm $\{u\}$.

2. **Definition of $H$-definitely self-adjoint systems.** In the following pages it will be assumed that the elements of the matrices $A(x)$ and $B(x)$ are real single-valued continuous functions on the finite interval $a\leqq x\leqq b$ and that the elements of $B(x)$ are not all identically zero on this interval. The elements of the matrices $M$ and $N$ are supposed to be real constants such that the $n\times 2n$ matrix $\|M_{ij}\ N_{ij}\|$ is of rank $n$. Moreover, because of its frequent occurrence, we write $\mathcal{L}[y]$ for the vector differential operator $y'-A(x)y$. The boundary value problem to be considered in this paper may then be written

$$\mathcal{L}[y] = \lambda B(x)y, \qquad s[y] \equiv My(a) + Ny(b) = 0. \tag{2.1}$$

The system adjoint to (2.1) is

$$\mathfrak{M}[z] = -\lambda zB(x), \qquad t[z] \equiv z(a)P + z(b)Q = 0, \tag{2.2}$$

where $\mathfrak{M}[z]$ is the adjoint differential operator $z'+zA(x)$, and $p\equiv(p_i)\equiv(P_{ij})$, $q\equiv(q_i)\equiv(Q_{ij})$, $(j=1,\cdots,n)$, are $n$ linearly independent solutions of the algebraic equations $Mp-Nq=0$.

According to the modified definition of Bliss [2] the system (2.1) is definitely self-adjoint with a matrix $T$, or simply definitely self-adjoint, whenever the following conditions are satisfied:

(i) The system is self-adjoint under the nonsingular real transformation $z=T(x)y$; that is, for arbitrary values of $\lambda$ a vector $y$ satisfies the differential equations (or boundary conditions) of (2.1) if and only if the associated vector $z=Ty$ satisfies the differential equations (or boundary conditions) of (2.2). The elements of $T(x)$ are supposed to be of class $C^1$ on the interval $ab$.

(ii) The matrix $S(x)=\tilde{T}(x)B(x)$ is symmetric on $ab$.

(iii) The quadratic form $uS(x)u$ is positive semi-definite on $ab$.

(iv) There exists no nonidentically vanishing solution $y$ of $\mathcal{L}[y]=0$, $s[y]=0$ such that $B(x)y(x)\equiv 0$ on $ab$.

The wording of hypothesis (iv) differs slightly from that used by Bliss [2, property (3), p. 414]. However, when (ii) and (iii) are satisfied the above hypothesis (iv) is readily seen to be equivalent to the property (3) of Bliss. For the present treatment the form (iv) is preferable.

It is to be remarked [1, p. 569] that a nonsingular matrix $T(x)$ whose elements are of class $C^1$ on $ab$ satisfies condition (i) if and only if

$$\begin{gathered} TA + \tilde{A}T + T' \equiv 0, \qquad TB + \tilde{B}T \equiv 0 \quad \text{on} \quad ab, \\ MT^{-1}(a)\tilde{M} = NT^{-1}(b)\tilde{N}. \end{gathered} \tag{2.3}$$

Consequently, whenever (ii) is also satisfied by $T$ we have

$$S \equiv \tilde{T}B \equiv \tilde{B}T \equiv -TB. \tag{2.4}$$

For $y$ a solution of (2.1) corresponding to a characteristic value $\lambda$, relation (1.2) becomes in matrix and vector notation

$$\int_a^b y\tilde{T}\mathcal{L}[y]\,dx = \lambda\int_a^b ySy\,dx. \tag{2.5}$$

Now the above hypothesis (iii) clearly implies a positive semi-definite character of the quadratic functional

$$\int_a^b ySy\,dx.$$

Indeed, (ii) and (iii) together imply that this functional is positive for all vectors $y$ whose components are continuous on $ab$ and such that $B(x)y(x)\not\equiv 0$ on this interval.

The quadratic functional upon which certain assumptions of definiteness are to be imposed in this paper is

$$H[y] \equiv \int_a^b y\tilde{T}\mathcal{L}[y]\,dx, \tag{2.6}$$

which appears as the left-hand member of (2.5).

For convenience, we shall denote by $L$ the linear vector space consisting of all vectors $y$ satisfying the following conditions: ($\alpha$) the components of $y$ are real-valued and of class $C^1$ on $ab$; ($\beta$) $s[y]=0$; ($\gamma$) there exists a corresponding vector $g(x)$ with real-valued continuous components such that $\mathcal{L}[y]=Bg$ on $ab$.

Instead of the above hypothesis (iii) we shall now assume the following condition:

(iii)′ The quadratic functional $H[y]$ is positive for arbitrary vectors $y$ of $L$ such that $B(x)y(x)\not\equiv 0$ on $ab$.

A system (2.1) which satisfies hypotheses (i), (ii), (iii)′ and (iv) will be termed *H-definitely self-adjoint with the matrix T*, or simply *H-definitely self-adjoint*; the prefix "*H-*" indicates that it is the functional $H[y]$ which possesses the definiteness property. Correspondingly, a system which is definitely self-adjoint as defined by Bliss [2] might be termed $S$-definitely self-adjoint. It is to be pointed out that in the treatment of definitely self-adjoint systems, as well as in the present discussion of $H$-definitely self-adjoint systems, the space $L$ occupies a central position.

Now a linear change of parameter in (2.1), replacing $\lambda$ by $\lambda+\lambda_0$, has the effect of substituting $\mathcal{L}[y]-\lambda_0 By$ for $\mathcal{L}[y]$. Hence it is to be emphasized that as far as the qualitative properties of (2.1) are concerned the hypothesis (iii)′

is no stronger than the assumption that there is some value of $\lambda_0$ such that the functional

$$H[y:\lambda_0] \equiv \int_a^b y\tilde{T}(\mathcal{L}[y] - \lambda_0 By)\, dx$$

satisfies the definiteness property of (iii)′.

In view of equations (2.3), a system (2.1) which is self-adjoint with a matrix $T$ is also self-adjoint with the matrix $-T$, $\tilde{T}$ or $-\tilde{T}$. In particular, if (2.1) is $H$-definitely self-adjoint with a matrix $T$ it is also $H$-definitely self-adjoint with the matrix $-\tilde{T}$. If hypotheses (i), (ii) and (iv) are satisfied by (2.1) with a matrix $T$ and the functional (2.6) is negative for arbitrary vectors $y$ of $L$ such that $By \not\equiv 0$, this functional can be replaced by one for which (iii)′ as stated is satisfied by using the transformation matrix $-T$ instead of $T$. Moreover, if (2.1) is $H$-definitely self-adjoint with a matrix $T$ then the adjoint system (2.2), written in the form (2.1), is $H$-definitely self-adjoint with the matrix $\tilde{T}^{-1}$. It may be readily verified that analogous results hold for definitely self-adjoint systems.

We shall denote by $L^2$ the linear subspace of $L$ consisting of all vectors $y$ with real components and satisfying a system $\mathcal{L}[y] = Bg$, $s[y] = 0$, where $g(x)$ is also a vector of the space $L$. Clearly each real characteristic solution of (2.1) belongs to $L^2$ as well as to $L$. In the subsequent discussion the space $L^2$ first occurs in Theorem 5.4.

3. **Properties of the quadratic functional** $H[y]$. If $u$ and $v$ are vectors whose components are of class $C^1$ on $ab$, let $H[u;v]$ denote the bilinear expression

$$H[u; v] = \int_a^b u\tilde{T}\mathcal{L}[v]\, dx.$$

In general $H[u; v] \neq H[v; u]$. However, we do have the following property.

LEMMA 3.1. *For a system* (2.1) *satisfying conditions* (i) *and* (ii) *the bilinear functional* $H[u; v]$ *is symmetric on the linear vector space* $L$.

For suppose that $u$ and $v$ belong to $L$, and that $\mathcal{L}[u] = Bg$, $\mathcal{L}[v] = Bh$. Then $w = Tv$ satisfies the system $\mathcal{M}[w] = -h\tilde{T}B$, $t[w] = 0$. By a familiar argument it then follows (see Bliss [1, p. 567]) that

$$\int_a^b wBg\, dx - \int_a^b hSu\, dx = wu \Big|_{x=a}^{x=b} = 0.$$

The result of the lemma is then immediate since

$$\int_a^b hSu\, dx = \int_a^b u\tilde{T}Bh\, dx = H[u; v],$$

$$\int_a^b wBg\, dx = \int_a^b v\tilde{T}Bg\, dx = H[v; u].$$

LEMMA 3.2. *Hypotheses* (ii) *and* (iii)′ *imply* $H[y] \geqq 0$ *on* $L$.

For consider an arbitrary vector $y$ of $L$. If $By \not\equiv 0$ on $ab$, then (iii)′ implies $H[y] > 0$. On the other hand, if $By \equiv 0$ on this interval the symmetry of $S$ insures

$$H[y] \equiv \int_a^b y\tilde{T}Bg\,dx = \int_a^b g\tilde{T}By\,dx = 0.$$

The following result is an immediate consequence of Lemma 3.1 and the linearity of $L$.

LEMMA 3.3. *If the system* (2.1) *satisfies* (i), (ii), *and* $H[y] \geqq 0$ *on* $L$, *then*

$$\{H[u; v]\}^2 \leqq H[u]H[v] \tag{3.1}$$

*for arbitrary vectors* $u$ *and* $v$ *of* $L$.

LEMMA 3.4. *If* (2.1) *satisfies hypotheses* (i) *and* (ii) *and* $y$, $y^*$ *are characteristic functions corresponding to distinct characteristic values* $\lambda$, $\lambda^*$, *then*

$$\int_a^b y^*Sy\,dx = 0, \qquad H[y^*; y] = 0. \tag{3.2}$$

The first equality of (3.2) follows by Theorem 8 of Bliss [1], and the second relation is then immediate since $y^*\tilde{T}\mathcal{L}[y] = \lambda y^*Sy$. It is to be remarked that this result is true quite independent of the reality of the characteristic values and characteristic functions involved.

4. **Preliminary results.** In this section we shall present some results for $H$-definitely self-adjoint systems which, although of individual significance, are preliminary to the rest of the paper.

THEOREM 4.1. *If* (2.1) *is H-definitely self-adjoint, then* $\lambda = 0$ *is not a characteristic value of this system.*

For suppose $\lambda = 0$ were a characteristic value for such a system, and denote by $y$ a corresponding real characteristic solution. The condition (iv) implies $By \not\equiv 0$ on $ab$, and as $y$ clearly belongs to $L$ it follows by (iii)′ that $H[y] > 0$. On the other hand, $H[y] = 0$ since $\mathcal{L}[y] \equiv 0$. Hence $\lambda = 0$ is not a characteristic value.

In connection with this theorem, we also have the following result.

THEOREM 4.2. *If* (2.1) *satisfies conditions* (i) *and* (ii), $H[y] \geqq 0$ *on* $L$, *and* $\lambda = 0$ *is not a characteristic value, then this system is H-definitely self-adjoint.*

If $\lambda = 0$ is not a characteristic value for (2.1), then clearly condition (iv) is satisfied. Moreover, suppose that $y$ is a particular vector of $L$ such that $H[y] = 0$. Then by Lemma 3.3 it follows that $H[y; v] = 0$ for arbitrary vectors $v$ of $L$. But for an arbitrary vector $g$ whose components are continuous there

exists, if $\lambda=0$ is not a characteristic value, a unique solution of $\mathcal{L}[v]=Bg$, $s[v]=0$; for this vector $v$, $H[y;v]=\int_a^b ySg\,dx$. Hence $0\equiv yS\equiv \tilde{T}By$, and therefore $By\equiv 0$ on $ab$. That is, if $y$ is a vector of $L$ for which $H[y]=0$ it must be true that $By\equiv 0$. Hence condition (iii)′ is also satisfied by such a system, and it is $H$-definitely self-adjoint.

As a consequence of Theorem 4.1, for an $H$-definitely self-adjoint system the functional $H[y]$ is afforded an alternate representation. Let $G(x,t)\equiv\|G_{ij}(x,t)\|$ denote the Green's matrix (see, for example, Bliss [1, pp. 577–581]) for the incompatible system

$$\mathcal{L}[y]=0,\qquad s[y]=0. \tag{4.1}$$

Then a vector $y$ belongs to $L$ if and only if it is of the form

$$y(x)=\int_a^b K(x,t)g(t)\,dt, \tag{4.2}$$

where $K(x,t)=G(x,t)B(t)$, and the components of $g$ are continuous on $ab$. We may then write

$$\begin{aligned} H[y] &= \int_a^b y\tilde{T}Bg\,dx = \int_a^b gSy\,dx \\ &= \int_a^b\int_a^b g(x)S(x)K(x,t)g(t)\,dx\,dt \\ &= \int_a^b\int_a^b g(x)K_1(x,t)g(t)\,dx\,dt, \end{aligned} \tag{4.3}$$

where $K_1(x,t)=S(x)K(x,t)=S(x)G(x,t)B(t)$. Similarly, if $u$ and $v$ are vectors belonging to $L$ and $\mathcal{L}[u]=Bg$, $\mathcal{L}[v]=Bh$, we have

$$H[u;v]=\int_a^b\int_a^b h(x)K_1(x,t)g(t)\,dx\,dt. \tag{4.3′}$$

Now corresponding to arbitrary vectors $g$, $h$ whose components are continuous there exist unique corresponding vectors $u$, $v$ of $L$ satisfying the above conditions. Since by Lemma 3.1 we have $H[u;v]=H[v;u]$ it then follows that

$$K_1(x,t)\equiv\tilde{K}_1(t,x), \tag{4.4}$$

that is,

$$S(x)K(x,t)\equiv\tilde{K}(t,x)S(t).$$

Indeed, in the proof of (4.4) we have used in addition to hypotheses (i), (ii) only the condition that $\lambda=0$ is not a characteristic value of (2.1). Relation (4.4) has been obtained by Bliss [1, p. 580], and it may be readily verified that his proof also uses only these conditions on the system (2.1).

For convenience in the presentation of the following two lemmas the functional (4.3) will be denoted by $J[g]$; similarly, the quantity (4.3′) will be written $J[h; g]$. The following result will be stated without proof, since it follows readily from well known properties of Lebesgue integrals.

LEMMA 4.1. *If $J[g] \geqq 0$ for all vectors $g$ whose components are continuous on $ab$, then this integral, taken in the sense of Lebesgue, is non-negative for all vectors $g$ whose components are of integrable square on this interval.*

Now denote by $L'$ the extension of $L$ obtained by considering the totality of vectors $y$ such that: ($\alpha'$) the components of $y$ are real-valued and absolutely continuous on $ab$; ($\beta'$) $s[y]=0$; ($\gamma'$) there exists a corresponding vector $g$ whose components are real-valued, of Lebesgue integrable square on $ab$, such that $\mathcal{L}[y]=Bg$ almost everywhere on this interval. In view of Lemma 4.1, and the fact that we still have $y(x)=\int_a^b K(x, t)g(t)dt$ for a vector $y$ of $L'$, it follows that the results of Lemmas 3.1, 3.2 and 3.3 remain valid for $y$, $u$ and $v$ vectors in $L'$. Moreover, since $\lambda=0$ is not a characteristic value for an $H$-definitely self-adjoint problem, it follows as in the proof of Theorem 4.2 that $H[y]=0$ for a vector $y$ of $L'$ if and only if $By\equiv 0$ on $ab$. That is, as far as the results previously established are concerned, in the definition of $H$-definitely self-adjoint systems one might without further restriction have used the vector space $L'$ instead of $L$. As a matter of fact, *this remark is valid for all the results obtained in the present paper*. Specifically, in this connection, it is to be pointed out that the results of Bliss used in §7 are valid for the space $L'$ instead of $L$.

Because of the special form of $K_1(x, t)$, and the fact that if the components of $b(x)$ are integrable on $ab$ then $y(x)=\int_a^b G(x, t)b(t)dt$ is a vector whose components are absolutely continuous, satisfies $\mathcal{L}[y]=b(x)$ almost everywhere on $ab$, and $s[y]=0$, all the preceding results may be proved for a much more general linear vector space than $L'$. In particular, they all hold for the space of vectors $y$ satisfying the above conditions ($\alpha'$), ($\beta'$), and the condition obtained by replacing in ($\gamma'$) the phrase "of Lebesgue integrable square" by "Lebesgue integrable." However, for a number of the subsequent results to remain valid, it is necessary to restrict the involved vectors to the space $L'$.

We shall denote by $K_{1ij}(x, x+)$ the limiting values of $K_{1ij}(x, t)$ as $t$ tends to $x$ through values greater than $x$, and write $K_1(x, x+)\equiv\|K_{1ij}(x, x+)\|$. The quantities $K_{1ij}(x, x-)$ and $K_1(x, x-)$ are defined in a corresponding fashion. Since $K_1(x, t)=S(x)G(x, t)B(t)$ it follows that the elements of $K_1$ have discontinuities at most along the line $x=t$. Moreover, if $K_1(x, t)$ is taken to be equal to $K_1(x, x+)$ along $x=t$, then the elements of the resulting matrix are continuous in $(x, t)$ on the region $R_1$: $x\leqq t\leqq b$, $a\leqq x\leqq b$. Similarly, if $K_1(x, t)$ is taken to be equal to $K_1(x, x-)$ along $x=t$, then the elements of the resulting matrix are continuous in $(x, t)$ on $R_2$: $a\leqq t\leqq x$; $a\leqq x\leqq b$.

LEMMA 4.2. *If the functional $J[g]$ defined by* (4.3) *be positive semi-definite for*

*arbitrary vectors g whose components are continuous on ab, then* $K_1(x, x-) = K_1(x, x+)$ *on* $a \leqq x \leqq b$; *that is, if* $K_1(x, t)$ *be taken as equal to this common value along the line* $x=t$, *then the elements of* $K_1$ *are continuous in* $(x, t)$ *on* $a \leqq x, t \leqq b$. *Moreover, the matrix* $K_1(x, x)$ *thus defined is symmetric and positive semi-definite on ab.*

For convenience, we write $J[g] = J_1[g] + J_2[g]$, where $J_1$ and $J_2$ are the integrals of the integrand of (4.3) taken over the above defined regions $R_1$ and $R_2$, respectively. The integrals $J_1[h; g]$, $J_2[h; g]$ are defined similarly. Now consider a point $(x_0, x_0)$ with $a < x_0 < b$. For an arbitrary constant vector $g_0$, denote by $g_k(x)$, $(k=1, 2, \cdots)$, the vector whose components are identically zero except on $|x - x_0| \leqq d_k$, where $d_k = c/k$ and $c$ is the smaller of the numbers $x_0 - a$, $b - x_0$, while on $x_0 - d_k \leqq x < x_0$ we define $g_k = (1/d_k)g_0$, and on $x_0 \leqq x \leqq x_0 + d_k$ we set $g_k = (-1/d_k)g_0$. Because of the continuity properties of the elements of $K_1$ as described above, it is readily calculated that $\lim_{k\to\infty} J_1[g_k] = 0 = \lim_{k\to\infty} J_2[g_k]$, and consequently, $\lim_{k\to\infty} J[g_k] = 0$. Now for a second arbitrary constant vector $h_0$, define $h_k(x) \equiv 0$ except on $x_0 - d_k \leqq x \leqq x_0 + d_k$, $h_k(x) \equiv (1/d_k)h_0$ on this subinterval. Clearly there exists a constant $\kappa$ such that $H[h_k] \leqq \kappa$, $(k=1, 2, \cdots)$. Moreover, in view of the positive semi-definite character of $J$ proved in Lemma 4.1, we then have $\{J[h_k; g_k]\}^2 \leqq J[h_k]\, J[g_k] \leqq \kappa J[g_k]$, $(k=1, 2, \cdots)$, and hence $\lim_{k\to\infty} J[h_k; g_k] = 0$. Again, using the continuity properties of the elements of $K_1$ described above, it is found that $\lim_{k\to\infty} J_1[h_k; g_k] = -h_0 K_1(x_0, x_0+)g_0$, and $\lim_{k\to\infty} J_2[h_k; g_k] = h_0 K_1(x_0, x_0-)g_0$. Thus for arbitrary constant vectors $h_0$, $g_0$ we have $h_0[K_1(x_0, x_0-) - K_1(x_0, x_0+)]g_0 = 0$, and consequently $K_1(x, x-) - K_1(x, x+) \equiv 0$ on $a < x < b$, whence it in turn follows that this relation is also true at the end values $a$ and $b$. In the following we shall write $K_1(x, x)$ for this common limiting value along the line $x=t$. Returning to the above defined sequence $\{h_k\}$, we see that for $a < x_0 < b$ we have $4h_0 K_1(x_0, x_0)h_0 = \lim_{k\to\infty} J[h_k] \geqq 0$ for arbitrary vectors $h_0$. Hence, the matrix $K_1(x, x)$ is positive semi-definite on $a < x < b$, and by continuity this property holds on the closed interval $ab$. The symmetry of this matrix follows from (4.4).

Clearly the above result applies to any positive semi-definite kernel matrix $K_1(x, t)$ such that $K_1(x, t) = \tilde{K}_1(t, x)$, and which possesses the continuity properties described immediately preceding the lemma.

THEOREM 4.3. *For an H-definitely self-adjoint system* (2.1) *the matrix* $B(x)$ *must satisfy the condition* $BB \equiv 0$ *on ab; in particular, the rank of* $B(x)$ *at any point of this interval cannot be greater than* $[n/2]$, *the largest integer not exceeding the value* $n/2$.

Since for an $H$-definitely self-adjoint system we have $K_1(x, t) = S(x)G(x, t)B(t)$, and as $G(x, x-) - G(x, x+) \equiv I$ on $ab$ (see, for example, [1, p. 578]), we have from Lemma 4.2 that $0 \equiv K_1(x, x-) - K_1(x, x+) \equiv S(x)B(x) \equiv \tilde{T}(x)B(x)B(x)$ on $ab$. As $T$ is nonsingular, it then follows that

$BB\equiv 0$ on this interval. Algebraically, it is readily seen that this condition implies that at each point of $ab$ the rank of $B$ cannot exceed $[n/2]$.

This result, which at first notice seems remarkable, first occurred to the author in considering the results which will be presented in §8. Indeed, because of this relatively strong condition imposed upon $B(x)$ by the $H$-definite self-adjoint character of (2.1), one might conclude that this class of boundary value problems is too restrictive to be of great significance. That this is not so, however, is borne out by the fact that this class of problems includes those of the type discussed in §10. Moreover, the additional results obtained in §7 concerning systems which are definitely self-adjoint in the sense of Bliss also show the significance of such problems.

5. **Properties of $H$-definitely self-adjoint systems.** We shall now proceed to establish some fundamental properties of systems (2.1) which are $H$-definitely self-adjoint.

THEOREM 5.1. *All the characteristic values of an H-definitely self-adjoint system are real, and the corresponding characteristic functions may be chosen real.*

Suppose $\lambda=\lambda_1+(-1)^{1/2}\lambda_2$, $(\lambda_2\neq 0)$, is a characteristic value of (2.1), and $y=u+(-1)^{1/2}v$ is a corresponding characteristic solution. Then $\bar{y}=u-(-1)^{1/2}v$ is a characteristic solution of this system corresponding to the complex conjugate value $\bar{\lambda}$ of the characteristic parameter. As $\lambda\neq\bar{\lambda}$, it follows from Lemma 3.4 that $H[\bar{y};y]=0$. Since

$$\mathcal{L}[u] = \lambda_1 Bu - \lambda_2 Bv, \qquad \mathcal{L}[v] = \lambda_2 Bu + \lambda_1 Bv, \qquad s[u] = 0 = s[v],$$

the vectors $u$, $v$ belong to $L$. Consequently, in view of Lemma 3.1, $H[\bar{y};\,y\,]=H[u]+H[v]$. It then follows from condition (iii)′ and Lemma 3.2 that $Bu\equiv 0$, $Bv\equiv 0$ on $ab$; that is, $u$ and $v$ are individually solutions of (2.1) for $\lambda=0$. It is then a consequence of Theorem 4.1 that $u\equiv 0$, $v\equiv 0$, which is a contradiction to the assumption that $y=u+(-1)^{1/2}v$ is a characteristic solution for the value $\lambda$. Hence all the characteristic values of an $H$-definitely self-adjoint system are real, and because of the reality of the coefficients of such a system the corresponding characteristic solutions may be chosen real.

In the future, when we speak of a characteristic solution of an $H$-definitely self-adjoint system, *it will be understood that this solution is real.*

THEOREM 5.2. *If $\lambda$ is a characteristic value of an H-definitely self-adjoint system, and y a corresponding characteristic solution, then $H[y]>0$ and $\int_a^b ySy\,dx$ has the sign of $\lambda$.*

Since, by Theorem 4.1, $\lambda=0$ is not a characteristic value, for a characteristic solution $y$ of (2.1) we have $By\not\equiv 0$ on $ab$, and hence $H[y]>0$ by (iii)′. The rest of the theorem is an immediate consequence of (2.5).

Let $Y(x,\lambda)$ be a matrix whose columns are $n$ linearly independent solutions of the differential equations of (2.1), and whose elements are perma-

nently convergent power series in $\lambda$. Such a matrix is determined, for example, by the initial condition $Y(a, \lambda) = I$. By definition, the *multiplicity* of a characteristic value of (2.1) is equal to its multiplicity as a zero of the characteristic determinant $|MY(a, \lambda)+NY(b, \lambda)|$, which is a permanently convergent power series in $\lambda$. The *index* of $\lambda$ as a characteristic value of (2.1) is equal to the number of corresponding linearly independent solutions of this system.

THEOREM 5.3. *For an H-definitely self-adjoint system* (2.1) *the index of a characteristic value is equal to its multiplicity.*

The proof is the same as that of Theorem 10 in Bliss [1], down to the last equation on page 572. On the assumption that the result of the theorem is not true, this equation states that there exists a characteristic solution $y$ of (2.1) such that $\int_a^b ySy\,dx = 0$. This, however, is impossible in view of the above Theorem 5.2.

THEOREM 5.4. *If the components of $f$ are continuous on $ab$, and the condition*

$$\int_a^b fSy\,dx = 0 \tag{5.1}$$

*is satisfied by every characteristic solution $y$ of an H-definitely self-adjoint system, then this relation is also satisfied by every vector $y$ of $L^2$.*

In view of the preceding theorem, the condition (5.1) for all characteristic solutions of (2.1) implies, as in the proof of Theorem 11 of Bliss [1], that the nonhomogeneous system

$$\mathcal{L}[y] = \lambda By + Bf, \qquad s[y] = 0, \tag{5.2}$$

has a solution $y(x, \lambda)$ of the form

$$y(x, \lambda) = u_0(x) + \lambda u_1(x) + \cdots + \lambda^\mu u_\mu(x) + \cdots; \tag{5.3}$$

the components of $u_\mu(x)$, $(\mu = 0, 1, \cdots)$, are of class $C^1$ on $ab$, and this series converges absolutely and uniformly on any region of the form $a \leqq x \leqq b$, $|\lambda| \leqq \rho$. Moreover, if we write $u_{-1}(x) = f(x)$ and $v_\mu(x) = T(x)u_\mu(x)$, $(\mu = -1, 0, 1, \cdots)$, then

$$\mathcal{L}[u_\mu] = Bu_{\mu-1}, \qquad s[u_\mu] = 0, \qquad \mu = 0, 1, \cdots, \tag{5.4}$$

$$\mathfrak{M}[v_\nu] = -v_{\nu-1}B, \qquad t[v_\nu] = 0, \qquad \nu = 0, 1, \cdots. \tag{5.5}$$

In view of the boundary conditions we also have

$$\int_a^b u_{\nu-1}Su_\mu\,dx = \int_a^b u_\nu Su_{\mu-1}\,dx, \qquad \mu, \nu = 0, 1, \cdots.$$

As

$$\int_a^b u_{\nu-1}Su_\mu\,dx = \int_a^b u_{\nu-1}\tilde{T}Bu_\mu\,dx = \int_a^b u_{\nu-1}\tilde{T}\mathcal{L}[u_{\mu+1}]\,dx,$$

$$\int^b u_\nu S u_{\mu-1}\,dx = \int_a^b u_\nu \tilde{T} B u_{\mu-1}\,dx = \int_a^b u_\nu \tilde{T}\mathcal{L}[u_\mu]\,dx,$$

we have

$$H[u_{\nu-1}; u_{\mu+1}] = H[u_\nu; u_\mu], \qquad \mu, \nu = 1, 2, \cdots.$$

By Lemma 3.1 it also follows that $H[u_\nu; u_\mu] = H[u_\mu; u_\nu]$. Now set

$$W_\mu = H[u_0; u_\mu], \qquad \mu = 0, 1, \cdots.$$

By the above relations we have

$$W_{\mu+\nu} = H[u_\nu; u_\mu], \qquad \mu, \nu = 0, 1, \cdots,$$

and it results from Lemma 3.3 that

$$(5.6) \qquad [W_{2\mu}]^2 = [W_{(\mu-1)+(\mu+1)}]^2 \leqq W_{2\mu-2}W_{2\mu+2}, \qquad \mu = 1, 2, \cdots.$$

Writing the differential equations of (5.2) in integral form, and employing the uniform convergence of (5.3) in a region of the form $a \leqq x \leqq b$, $|\lambda| \leqq \rho$, it follows that $\mathcal{L}[y]$ is a permanently convergent power series in $\lambda$ given by

$$(5.7) \qquad \begin{aligned} &\mathcal{L}[u_0] + \lambda\mathcal{L}[u_1] + \cdots + \lambda^\mu \mathcal{L}[u_\mu] + \cdots \\ &\qquad = Bf + \lambda B u_0 + \cdots + \lambda^\mu B u_{\mu-1} + \cdots. \end{aligned}$$

From its specific form, it is seen that the series (5.7) has convergence properties of the sort indicated above for the series (5.3). Consequently, the series

$$(5.8) \qquad W_0 + \lambda W_1 + \lambda^2 W_2 + \cdots, \qquad W_0 + \lambda^2 W_2 + \cdots,$$

the first of which is equal to $H[u_0; y]$, are permanently convergent power series in $\lambda$.

If $W_2 \neq 0$ it follows from (5.6) that $W_{2\mu} \neq 0$, $(\mu = 1, 2, \cdots)$, and the second series of (5.8) is seen to diverge for $\lambda = (W_2/W_4)^{1/2}$. Hence the permanent convergence of this series is possible only if $0 = W_2 = H[u_1]$. Condition (iii)$'$ then implies that $Bu_1 \equiv 0$ on $ab$. Moreover, by Lemma 3.3 we have $H[y; u_1] = 0$ for arbitrary vectors $y$ of $L$. As $\mathcal{L}[u_1] = Bu_0$, we may also state this condition as

$$(5.9) \qquad 0 = \int_a^b y\tilde{T}\mathcal{L}[u_1]\,dx = \int_a^b ySu_0\,dx$$

for arbitrary vectors $y$ of $L$. In particular, for $y = u_0$ we have

$$(5.10) \qquad \int_a^b u_0 S u_0\,dx = 0.$$

Now suppose $y$ is any vector belonging to the space $L^2$, and $g(x)$ is a vector of $L$ such that $\mathcal{L}[y] = Bg$. By Lemma 3.1 we then have

$$0 = H[y; u_0] - H[u_0; y] = \int_a^b ySf\,dx - \int_a^b u_0Sg\,dx$$
$$= \int_a^b ySf\,dx,$$

in view of (5.9) and the fact that $g$ is a vector of $L$. This completes the proof of Theorem 5.4.

The above result for $H$-definitely self-adjoint systems is somewhat weaker than the corresponding result for definitely self-adjoint systems (see Bliss [2, Theorem 2.3]). Formally, this is true because the permanent convergence of the second series of (5.8) does not imply that the constant term $W_0$ of this series is equal to zero; the failure to obtain this latter result is in turn a consequence of the fact that we do not have an inequality of the form (5.6) for $\mu=0$. If the convergence of the second series of (5.8) were to imply the vanishing of $W_0$, by the argument used above we could proceed to show that the hypotheses of the above theorem imply the relation (5.1) for arbitrary vectors of the space $L$ instead of merely for the vectors belonging to $L^2$. That the result of the above theorem cannot in general be thus strengthened, however, is shown by the following example.

Consider the system

$$\begin{aligned} &y_1' = 0, && y_2' = -\lambda b(x)y_1, \\ &y_1(0) - y_2(0) = 0, && y_1(1) + y_2(1) = 0, \end{aligned} \tag{5.11}$$

where $b(x)$ is a continuous function not identically zero on $0 \leqq x \leqq 1$, and such that

$$\int_0^1 b(x)\,dx = 0. \tag{5.12}$$

It may readily be verified that this system is $H$-definitely self-adjoint with the matrix

$$T \equiv \begin{Vmatrix} 0 & 1 \\ -1 & 0 \end{Vmatrix}, \tag{5.13}$$

and, moreover, this system has no characteristic values. For this system, therefore, the condition of Theorem 5.4 that (5.1) hold for every characteristic solution imposes no additional restriction on a vector $f \equiv (f_1(x), f_2(x))$ with continuous components. Now

$$y_1(x) \equiv 1, \qquad y_2(x) = 1 - \left(2 \Big/ \int_0^1 b^2(t)\,dt\right)\int_0^x b^2(t)\,dt$$

is a vector of the space $L$ for this problem, and for this particular $y$ we have

$$\int_0^1 fSy\,dx = \int_0^1 f_1(x)b(x)y_1(x)\,dx = \int_0^1 f_1(x)b(x)\,dx.$$

For certain continuous functions $f_1(x)$, in particular, for $f_1(x)=b(x)$, this expression is different from zero. Thus we see that in the statement of Theorem 5.4 the phrase "every vector $y$ of $L^2$" cannot in general be replaced by "every vector $y$ of $L$."

We shall now proceed to establish as corollaries to the above theorem certain related results.

COROLLARY 1. *If the system* (2.1) *is H-definitely self-adjoint and $f(x)$ is a vector of the corresponding space L for which condition* (5.1) *is satisfied by every characteristic solution y of the system, then this condition is also satisfied by every vector y of L; in particular,* $\int_a^b fSf\,dx=0$.

Let $g(x)$ be a vector with continuous components such that $\mathfrak{L}[f]=Bg$, $s[f]=0$. For a characteristic solution $y$ corresponding to a characteristic value $\lambda$ we have

$$\int_a^b ySg\,dx = H[y;f] = H[f;y] = \lambda\int_a^b fSy\,dx,$$

and thus the condition that $f$ satisfies (5.1) with every characteristic solution $y$ implies that the vector $g$ satisfies a like condition. It then follows from Theorem 5.4 that $\int_a^b y^*Sg\,dx=0$ for arbitrary vectors $y^*$ of $L^2$. Now for an arbitrary vector $y$ of $L$ let $y^*$ denote the vector of $L^2$ such that $\mathfrak{L}[y^*]=By$, $s[y^*]=0$. Then

$$0 = \int_a^b y^*Sg\,dx = H[y^*;f] = H[f;y^*] = \int_a^b fSy\,dx,$$

so that the conditions of the corollary imply (5.1) for arbitrary vectors $y$ of $L$. Since $f(x)$ belongs to $L$, we have, in particular, $\int_a^b fSf\,dx=0$.

COROLLARY 2. *If the system* (2.1) *is H-definitely self-adjoint and $f(x)$ is a vector of the corresponding space $L^2$ for which condition* (5.1) *is satisfied by every characteristic solution y of the system, then $B(x)f(x)\equiv 0$ on the interval ab.*

Let $g(x)$ be a vector of $L$ such that $\mathfrak{L}[f]=Bg$, $s[f]=0$. Then by an argument similar to that used in the proof of Corollary 1 we have $\int_a^b ySg\,dx=0$ for every characteristic solution $y$ and hence, by Theorem 5.4, this condition also holds for arbitrary vectors $y$ of $L^2$. In particular, for $y=f$ we have

$$0 = \int_a^b fSg\,dx = H[f],$$

and in view of (iii)′ we have $Bf\equiv 0$ on $ab$.

COROLLARY 3. *If for an H-definitely self-adjoint system the condition* $B(x)y(x)\equiv 0$ *on ab holds for a vector y of L if and only if* $y(x)\equiv 0$ *on this interval, then if the components of* $f(x)$ *are continuous and condition* (5.1) *is satisfied by every characteristic solution of the system it follows that* $B(x)f(x)\equiv 0$ *on the interval ab.*

In the proof of Theorem 5.4 it was established that the vector $u_1$ of $L$ defined by (5.3) satisfies $Bu_1\equiv 0$ on $ab$. Under the strengthened hypotheses of the corollary we consequently have $u_1\equiv 0$, and it then follows from (5.4) for $\mu=1$ that $Bu_0\equiv 0$ on $ab$. As $u_0$ is also a vector of $L$ it in turn results that $u_0\equiv 0$, and hence $B(x)f(x)\equiv 0$ on $ab$ by equation (5.4) for $\mu=0$.

For an $H$-definitely self-adjoint system the additional hypothesis of the above corollary is clearly equivalent to the following: $H[y]>0$ *for every non-identically vanishing vector y of L.*

6. **Existence of characteristic values.** In general an $H$-definitely self-adjoint system (2.1) does not possess an infinity of characteristic values. In particular, the example (5.11) of the preceding section illustrates the possibility that such a system may have no characteristic values. It is also easy to construct examples of such systems that have only a finite number of characteristic values. We shall, therefore, consider in this section the possible character of the totality of characteristic values of an $H$-definitely self-adjoint system.

Since for such a system the characteristic values are the zeros of a permanently convergent power series, and the index of each characteristic value is equal to its multiplicity, it follows that there can exist at most a denumerable infinity of characteristic values. Let $\{y_\mu, \lambda_\mu\}$, $(\mu=1, 2, \cdots)$, denote a maximal set of linearly independent characteristic solutions and corresponding characteristic values, the former chosen orthonormal in the sense that

$$(6.1)\qquad \int_a^b y_\mu S y_\nu\, dx = \delta_{\mu\nu}\frac{|\lambda_\mu|}{\lambda_\mu}, \qquad \mu, \nu = 1, 2, \cdots,$$

where $\delta_{\mu\nu}=0$ if $\mu\neq\nu$, $\delta_{\mu\mu}=1$. Such a choice is possible in view of Theorem 5.2.

THEOREM 6.1. *A necessary and sufficient condition that an H-definitely self-adjoint system have at least k linearly independent characteristic solutions is that the quadratic functional* $H[y]$ *be positive definite on a linear subspace of* $L^2$ *of dimension k; that is, that there exist vectors* $f_\mu(x)$, $(\mu=1, \cdots, k)$, *of* $L^2$ *such that for arbitrary constants* $(d_1, \cdots, d_k)\neq(0, \cdots, 0)$ *the vector* $f(x)=f_1(x)d_1+\cdots+f_k(x)d_k$ *renders* $H[f]>0$.

For suppose that $y_\mu(x)$, $(\mu=1, \cdots, k)$, are linearly independent characteristic solutions of such a system, and that these solutions are chosen orthonormal in the sense of (6.1). If $\lambda_\mu$ denote the characteristic value corresponding to $y_\mu$, then for $f_\mu=y_\mu$ and arbitrary constants $(d_1, \cdots, d_k)\neq(0, \cdots, 0)$ we

have for each vector $f=f_1d_1+\cdots+f_kd_k$ that $H[f]=|\lambda_1|d_1^2+\cdots+|\lambda_k|d_k^2>0$. Hence the condition of the theorem is necessary.

In order to prove the sufficiency of the theorem, suppose that there exists a linear subspace of $L^2$ determined by vectors $f_1,\cdots,f_k$ on which $H[y]$ is positive definite, while the system (2.1) has fewer than $k$ linearly independent characteristic solutions. It would then follow that there exists a set of constants $d_1,\cdots,d_k$ not all zero and such that the vector $f=f_1d_1+\cdots+f_kd_k$ satisfies equation (5.1) with every characteristic solution $y$. Since $f$ belongs to $L^2$, it is then a consequence of Corollary 2 to Theorem 5.4 that $Bf\equiv 0$, and hence $H[f]=0$ also, contrary to the assumption of the positive definite character of $H[y]$ on the linear subspace of $L^2$ determined by $f_1,\cdots,f_k$. Hence the condition is also sufficient.

We shall now give a particular sufficient condition for an $H$-definitely self-adjoint system to have an infinity of characteristic values. This condition has application for the special boundary value problem of §10. Suppose that the matrices $A(x)$ and $B(x)$ satisfy the following condition.

(v) There is a subinterval $a_1b_1$, $a<a_1<b_1<b$, of $ab$ such that if $a_1'$, $b_1'$ are arbitrary values satisfying $a_1\leqq a_1'<b_1'\leqq b_1$, then there exists a vector $g$ of $L$ and associated $y$ of $L^2$ satisfying $\mathcal{L}[y]=Bg$, $By\not\equiv 0$ on $a_1'b_1'$, whereas $y\equiv 0$ outside the given interval $a_1'b_1'$.

Theorem 6.2. *If an $H$-definitely self-adjoint system satisfies condition* (v), *then this system has infinitely many characteristic values.*

For consider an interval $a_1b_1$ on which the condition (v) is satisfied, and for a given integer $k$ divide $a_1b_1$ into $k$ non-overlapping subintervals $\Delta_1,\cdots,\Delta_k$. Let $y=f_\mu$ denote a vector of $L^2$ satisfying the conditions of (v) relative to $\Delta_\mu$, $(\mu=1,\cdots,k)$. Since $Bf_\mu\not\equiv 0$ on $\Delta_\mu$, and $f_\mu\equiv 0$ outside this interval, it follows readily that $H[f_\mu]>0$, $H[f_\mu;f_\nu]=0$ for $\mu\neq\nu$, $(\mu;\nu=1,\cdots,k)$. Consequently, for each $f=f_1d_1+\cdots+f_kd_k$ we have $H[f]=H[f_1]d_1^2+\cdots+H[f_k]d_k^2$, and by Theorem 6.1 the system (2.1) has at least $k$ linearly independent characteristic solutions. Since $k$ may be chosen arbitrarily, such a system has infinitely many characteristic values.

Corresponding to a vector $f$ we shall denote by $c_\mu[f]$ the Fourier coefficients

$$c_\mu[f]=\frac{|\lambda_\mu|}{\lambda_\mu}\int_a^b fSy_\mu\,dx,\qquad \mu=1,2,\cdots.$$

Clearly these coefficients are well-defined for a vector $f$ whose components are merely integrable on $ab$.

Lemma 6.1. *If* $\{y_\mu,\lambda_\mu\}$, $(\mu=1,2,\cdots)$, *denote a maximal set of linearly independent characteristic solutions and corresponding characteristic values for an $H$-definitely self-adjoint system* (2.1), *the former orthonormal in the sense of* (6.1), *then for an arbitrary vector $f$ of $L$,*

$$\sum_{\mu} |\lambda_\mu| c_\mu^2[f] \leq H[f]. \tag{6.2}$$

If $f$ belongs to $L$, then for arbitrary integers $k$ the vector $f-\sum_{\mu\leq k} y_\mu(x)c_\mu[f]$ is also in $L$, and

$$0 \leq H\left[f - \sum_{\mu\leq k} y_\mu c_\mu\right] = H[f] - \sum_{\mu\leq k} |\lambda_\mu| c_\mu^2[f].$$

7. **Definitely self-adjoint systems.** In this section we shall consider systems (2.1) that are definitely self-adjoint in the sense of Bliss. A maximal set of linearly independent characteristic solutions and associated characteristic values for such a system will again be denoted by $\{y_\mu(x), \lambda_\mu\}$, $(\mu = 1, 2, \cdots)$; moreover, we shall assume that the former are chosen orthonormal in the sense that

$$\int_a^b y_\mu S y_\nu \, dx = \delta_{\mu\nu}, \qquad \mu, \nu = 1, 2, \cdots. \tag{7.1}$$

We also write

$$e_\mu[f] = \int_a^b f S y_\mu \, dx, \qquad \mu = 1, 2, \cdots, \tag{7.2}$$

for the Fourier coefficients of a vector $f(x)$. It then follows from Theorem 3.1 of Bliss [2] that for an arbitrary vector $f$ of $L$ the series

$$\phi(x) = \sum_{\mu} y_\mu(x) e_\mu[f] \tag{7.3}$$

converges absolutely and uniformly on $ab$; moreover, $B(x)[f(x)-\phi(x)] \equiv 0$ on this interval.

THEOREM 7.1. *If (2.1) is definitely self-adjoint, then for arbitrary vectors $f$ of $L$ we have*

$$H[f] = \sum_{\mu} \lambda_\mu e_\mu^2[f], \qquad \int_a^b f S f \, dx = \sum_{\mu} e_\mu^2[f]. \tag{7.4}$$

The uniform convergence of the series (7.3) permits the evaluation of $H[f]$ as

$$\begin{aligned} H[f] = \int_a^b f S g \, dx &= \int_a^b g S f \, dx \\ &= \int_a^b g S \phi \, dx \\ &= \sum_{\mu} e_\mu[g] e_\mu[f] \\ &= \sum_{\mu} \lambda_\mu e_\mu^2[f], \end{aligned}$$

where $g$ is a vector such that $\mathcal{L}[f]=Bg$; the last relation above is a consesequence of the readily established equality $e_\mu[g]=\lambda_\mu e_\mu[f]$. Similarly

$$\int_a^b fSf\,dx = \int_a^b fS\phi\,dx = \sum_\mu e_\mu^2[f].$$

THEOREM 7.2. *Suppose that* (2.1) *is definitely self-adjoint and that the characteristic values of this system are bounded below; moreover, let the set* $\{y_\mu(x), \lambda_\mu\}$ *be so ordered that* $\lambda_1 \leqq \lambda_2 \leqq \cdots$. *If* $C_1$ *denote the totality of vectors* $f$ *of* $L$ *satisfying* $\int_a^b fSf\,dx=1$ *and* $C_1$ *is nonvacuous, then* $\lambda_1$ *is the minimum of* $H[f]$ *in this class; moreover, this minimum is attained by a particular* $f$ *of* $C_1$ *if and only if* $f=Y_1(x)+\Phi_1(x)$, *where* $Y_1$ *is a characteristic solution for* $\lambda=\lambda_1$ *and* $\Phi_1$ *is a vector of* $L$ *such that* $B\Phi_1\equiv 0$. *In general, if* $\lambda_1, \cdots, \lambda_{m-1}$ *exist, denote by* $C_m$ *the totality of vectors* $f$ *of* $L$ *satisfying*

$$\int_a^b fSf\,dx = 1, \qquad e_\mu[f] \equiv \int_a^b fSy_\mu\,dx = 0, \qquad \mu = 1, \cdots, m-1.$$

*If this class is nonvacuous, then* $\lambda_m$ *exists and is the minimum of* $H[f]$ *in* $C_m$; *moreover, this minimum is attained by a particular* $f$ *of* $C_m$ *if and only if* $f=Y_m(x)+\Phi_m(x)$, *where* $Y_m$ *is a characteristic solution for* $\lambda=\lambda_m$ *and* $\Phi_m$ *is a vector of* $L$ *such that* $B\Phi_m\equiv 0$ *on* $ab$.

The relations (7.4) clearly imply $H[f]\geqq\lambda_1$ in $C_1$ whenever this class is nonvacuous; furthermore, if $\lambda_1=\lambda_2=\cdots=\lambda_q<\lambda_{q+1}$, then the equality sign holds if and only if $e_\mu[f]=0$, $(\mu=q+1, \cdots)$. If $Y_1(x)=y_1(x)e_1[f]+\cdots+y_q(x)e_q[f]$, then $\Phi_1(x)=f(x)-Y_1(x)$ belongs to $L$ and $e_\mu[\Phi_1]=0$, $(\mu=1, 2, \cdots)$. Hence by the second equation of (7.4), and the definiteness of $S$ we have $B\Phi_1\equiv 0$ on $ab$ (see also Bliss [2, Corollary 2.2]). In general, if $\lambda_1, \cdots, \lambda_{m-1}$ exist and the class $C_m$ is nonvacuous, it again follows from (7.4) that $\lambda_m$ must exist and $H[f]\geqq\lambda_m$ in this class. Moreover, if $\lambda_m=\lambda_{m+1}=\cdots=\lambda_{m+p}<\lambda_{m+p+1}$, then the equality sign holds if and only if $e_\mu[f]=0$ for $\mu>m+p$. If we now define $Y_m(x)=y_m(x)e_m[f]+\cdots+y_{m+p}(x)e_{m+p}[f]$, then $\Phi_m=f-Y_m$ is a vector of $L$ such that $e_\mu[\Phi_m]=0$, $(\mu=1, 2, \cdots)$. Then, as above, it follows that $B\Phi_m\equiv 0$ on $ab$.

THEOREM 7.3. *If* (2.1) *is definitely self-adjoint and its characteristic values are either bounded below or above, then without loss of generality this system may be taken to be H-definitely self-adjoint; moreover, in this case* $BB\equiv 0$ *on* $ab$, *and the rank of* $B(x)$ *does not exceed* $[n/2]$ *at any point of this interval.*

Suppose that the characteristic values are bounded below, and let $\lambda_0$ be a number less than the smallest characteristic value, $\lambda_1$. Then for an arbitrary vector $f$ of $L$ the functional

$$H[f:\lambda_0] = \int_a^b f\tilde{T}(\mathcal{L}[f] - \lambda_0 Bf)\,dx = H[f] - \lambda_0\int_a^b fSf\,dx$$

may be written, in view of (7.4), as

$$H[f:\lambda_0] = \sum_\mu (\lambda_\mu - \lambda_0)e_\mu^2[f].$$

Consequently, for $f$ belonging to $L$ we have $H[f:\lambda_0] \geqq 0$, and the equality sign holds if and only if $e_\mu[f]=0$, $(\mu=1, 2, \cdots)$. For a definitely self-adjoint system, however, this condition implies $Bf\equiv 0$ on $ab$ in view of the second equation of (7.4) (see also Bliss [2, Corollary 2.2]). The replacement of $H[f]$ by $H[f:\lambda_0]$ is equivalent to a linear change of parameter in the boundary value problem (2.1). Hence for a definitely self-adjoint problem whose characteristic values are bounded below we may without loss of generality assume that the functional $H[y]$ satisfies the definiteness property (iii)$'$; that is, that the system is $H$-definitely self-adjoint. By Theorem 4.3 it then follows that $BB\equiv 0$ and the rank of this matrix does not exceed $[n/2]$ on $ab$.

In case the characteristic values of (2.1) are bounded above then the replacement of $\lambda$ by $-\lambda$, or the equivalent replacement of $B(x)$ by $-B(x)$, transforms the given system into one whose characteristic values are bounded below. The original system being definitely self-adjoint with $T$ implies that the new system is definitely self-adjoint with $-T$. Hence, by a linear change of parameter and the replacement of $T$ by $-T$ the given system is reducible to one which is $H$-definitely self-adjoint and the results of the theorem follow from the preceding case.

In this connection, it is worthwhile to point out that certain specific representations of "equivalent" boundary value problems may have individual advantages. For example, consider the boundary value problem $y''+\lambda y=0$, $y(0)=0=y(\pi)$. A maximal set of linearly independent characteristic functions and associated characteristic values is $\{\sin nx, n^2\}$, $(n=1, 2, \cdots)$. If we write this problem as $y_1'=y_2$, $y_2'=-\lambda y_1$, $y_1(0)=0=y_1(\pi)$, then this system is definitely self-adjoint and also $H$-definitely self-adjoint with the matrix (5.13). On the other hand, $y_1'=\rho y_2$, $y_2'=-\rho y_1$, $y_1(0)=0=y_1(\pi)$ is "equivalent" to the given problem by setting $\lambda=\rho^2$. This latter system is definitely self-adjoint with (5.13), but is clearly not $H$-definitely self-adjoint with this or any other matrix $T$ since the corresponding matrix $B(x)$ is nonsingular.

The following result is an immediate consequence of Theorem 7.3.

THEOREM 7.4. *If (2.1) is definitely self-adjoint and $B(x)B(x)\not\equiv 0$ on $ab$, then this system has infinitely many negative, and also infinitely many positive, characteristic values.*

This theorem contains as a very special case the result of Theorem 4.3 of Reid [10]. To see this, suppose that $B(x)$ has constant rank $n-m$ on $ab$ and denote by $\pi_i=\pi_{i\alpha}(x)$, $(\alpha=1, \cdots, m)$, linearly independent solutions of the equations $B(x)\pi=0$. From (2.4) it follows that $\tilde{T}^{-1}\tilde{B}\equiv BT^{-1}$, and hence the rank of $\|\pi_{i\alpha}(x)\, T_{ki}^{-1}(x)B_{jk}(x)\|$ is the same as the rank of $\|\,\pi_{i\alpha}(x)\, B_{ik}(x)T_{kj}^{-1}(x)\|$,

which in turn is equal to the rank of $\|\pi_{i\alpha}(x)\ B_{ij}(x)\|$. Now the rank of this latter matrix is clearly equal to $m$ if $BB\equiv 0$, and hence the hypotheses of Theorem 4.3 of [10], which demand that the rank of this matrix exceed $m$ at some point $x_0$ of $ab$, require that $BB\not\equiv 0$ on $ab$. It is also to be noted that in the above referred to theorem of [10] it was not proved that the system had infinitely many characteristic values of each algebraic sign, but simply that the system had infinitely many characteristic values under the stronger hypotheses there stated.

In view of relations (7.4) we have the following result.

THEOREM 7.5. *A definitely self-adjoint system* (2.1) *has at least k characteristic values if and only if* $\int_a^b fSf\,dx$ *is positive definite on a linear subspace of L of dimension k. Moreover, for a given constant* $\lambda_0$ *such a system has at least k characteristic values greater* [*less*] *than* $\lambda_0$ *if and only if the functional* $H[f:\lambda_0]$ *is positive* [*negative*] *definite on a linear subspace of L of dimension k.*

In the case of a definitely self-adjoint system for which the matrix $B(x)$ has constant rank on $ab$ the first part of this theorem was deduced by Reid [10, Theorem 4.1] from known results for an auxiliary problem associated with the calculus of variations. Analogues of the above theorem for $H$-definitely self-adjoint systems are contained in Theorem 6.1 and Theorem 9.3.

8. **A special definitely self-adjoint system.** Suppose now that the boundary value problem (2.1) satisfies conditions (i), (ii) and (iv) of §2 with a matrix $T(x)$. In this section we shall consider the associated system

$$\mathcal{L}[y] = \lambda B_1(x)y, \qquad s[y] = 0, \tag{8.1}$$

where $B_1(x)=B(x)\tilde{T}(x)B(x)=B(x)S(x)$. This problem is seen to be definitely self-adjoint with the same matrix $T(x)$. In the first place, in order to show that (8.1) is self-adjoint with $T$ it remains only to show that $TB_1+\tilde{B}_1T\equiv 0$, and this is true since $TB\tilde{T}B+\tilde{B}T\tilde{B}T\equiv(TB+\tilde{B}T)S\equiv 0$ by (2.3). If we set $S_1(x)=\tilde{T}(x)B_1(x)=S(x)S(x)$, clearly conditions (ii) and (iii) are satisfied by $S_1$. Finally, since $B_1y\equiv 0$ implies $y\tilde{T}B_1y\equiv ySSy\equiv 0$, and hence $Sy\equiv 0$ and $By\equiv 0$, condition (iv) for (2.1) implies the corresponding condition for (8.1).

Since a definitely self-adjoint problem has at most a denumerable infinity of characteristic values, for the consideration of (8.1) one may assume without loss of generality that $\lambda=0$ is not a characteristic value of this system. If this condition is not true for the problem as written, it is attainable by a linear change of parameter. *We shall make this assumption in the following discussion.*

If $y$ is a characteristic solution of (8.1) corresponding to a value $\lambda$, set $u(x)=S(x)y(x)$. In view of condition (iv) for (8.1) we have $u\not\equiv 0$ on $ab$. Then $\mathcal{L}[y]=\lambda Bu$, $s[y]=0$, and if $G(x,t)$ denotes the Green's matrix for the incompatible system $\mathcal{L}[y]=0$, $s[y]=0$, we have

$$y(x) = \lambda \int_a^b K(x,t)u(t)\,dt,$$

where, as in §4, $K(x, t) = G(x, t)B(t)$. In particular, it then follows that $u(x)$ is a characteristic solution, for this same value of $\lambda$, of the linear vector integral equation

$$u(x) = \lambda \int_a^b K_1(x, t)u(t)\,dt, \tag{8.2}$$

where, again as in §4, we have written $K_1(x, t) = S(x)K(x, t)$. It also follows from the comment after equation (4.4) that $K_1(x, t) = \tilde{K}_1(t, x)$, and hence (8.2) is of the type covered by the Hilbert-Schmidt theory of linear integral equations. Conversely, if $u$ is a characteristic solution of (8.2) for a value $\lambda$, and $y$ is defined as the corresponding unique solution of $\mathcal{L}[y] = \lambda Bu$, $s[y] = 0$, it follows that $u(x) = S(x)y(x)$ and $y$ is a characteristic solution of (8.1) for the same value of $\lambda$. Hence, there is complete equivalence between the boundary value problem (8.1) and the integral equation (8.2).

We shall denote by $\{y_\sigma, \Lambda_\sigma\}$, $(\sigma = 1, 2, \cdots)$, a maximal set of linearly independent characteristic solutions and corresponding characteristic values of (8.1), the former chosen orthonormal in the sense that

$$\int_a^b \tilde{y}_\sigma S_1 y_\tau\,dx = \delta_{\sigma\tau}, \qquad \sigma, \tau = 1, 2, \cdots. \tag{8.3}$$

Correspondingly, $\{u_\sigma = Sy_\sigma, \Lambda_\sigma\}$ is a maximal set of linearly independent characteristic solutions and corresponding characteristic values of (8.2) satisfying $\int_a^b \tilde{u}_\sigma u_\tau\,dx = \delta_{\sigma\tau}$.

Finally, if $g(x)$ is a vector whose components are continuous (or of Lebesgue integrable square) on $ab$, and if $f$ is defined by the system $\mathcal{L}[f] = Bg$, $s[f] = 0$, it follows from (4.3) that

$$H[f] = \int_a^b \int_a^b \tilde{g}(x)K_1(x, t)g(t)\,dx\,dt.$$

It is to be emphasized that the above defined vector $f$ belongs to the linear vector space $L$ for the problem (2.1), but not necessarily to the corresponding space $L_1$ for the problem (8.1), since this latter space contains vectors $f$ which satisfy with associated vectors $g$ the differential system

$$\mathcal{L}[f] = BSg, \qquad s[f] = 0. \tag{8.4}$$

In case the matrix $B$ is nonsingular on $ab$ the space $L_1$ is seen to be identical with $L$. However, since in general $L_1$ is a subspace of $L$ and $B_1y \equiv 0$ on $ab$ if and only if $By \equiv 0$ on this interval, the condition that (2.1) be $H$-definitely self-adjoint clearly implies that (8.1) is also $H$-definitely self-adjoint.

The results of Bliss [2], and those of the preceding section, give properties of the particular definitely self-adjoint system (8.1) on the space $L_1$. We wish, however, to go further and obtain properties of this system on the space $L$ corresponding to the boundary value problem (2.1). As pointed out above,

one may always by a linear change of parameter, replacing $\lambda$ by a suitable $\lambda+\lambda_0$, insure for (8.1) that $\lambda=0$ is not a characteristic value of this system. Now this change of parameter is equivalent to replacing $A(x)$ by

$$A(x)+\lambda_0 B(x)S(x).$$

Before proceeding further, it is to be emphasized that the space $L$, *as defined in* §2, is invariant under this operation. This results from the fact that for a given vector $y$ whose components are of class $C^1$ the vector differential equation $y'-Ay=Bg$ is equivalent to the equation $y'-Ay-\lambda_0 BSy=Bg_1$ by the transformation $g_1=g-\lambda_0 Sy$.

Corresponding to a given vector $f$, we set

$$d_\sigma[f] = \int_a^b fS_1 y_\sigma\, dx, \qquad \delta_\sigma[f] = \int_a^b fSy_\sigma\, dx = \int_a^b fu_\sigma\, dx, \qquad \sigma = 1, 2, \cdots;$$

clearly these coefficients are well-defined if the components of $f$ are integrable on $ab$. Since the vectors $u_\sigma$ are orthonormal, the following result is an immediate consequence of Bessel's inequality.

LEMMA 8.1. *If the components of $g(x)$ are of integrable square on $ab$, then the series $\sum_\sigma \delta_\sigma^2[g]$ converges and*

$$\sum_\sigma \delta_\sigma^2[g] \leqq \int_a^b gg\, dx. \tag{8.5}$$

LEMMA 8.2. *The series*

$$\sum_\sigma \left\{\frac{y_{i\sigma}(x)}{\Lambda_\sigma}\right\}^2, \qquad i = 1, 2, \cdots, n, \tag{8.6}$$

*converge on $ab$; moreover, if $\sum_\sigma \delta_\sigma^2 < +\infty$, the vector series*

$$\sum_\sigma \frac{y_\sigma(x)}{\Lambda_\sigma}\delta_\sigma \tag{8.7}$$

*converges absolutely and uniformly on this interval.*

Since

$$\frac{y_\sigma(x)}{\Lambda_\sigma} = \int_a^b K(x, t)u_\sigma(t)\, dt,$$

it follows that for a fixed value of $x$ each row of $K(x, t)$ is a vector satisfying the conditions of Lemma 8.1. Hence the series (8.6) converges; moreover, in view of (8.5), there clearly exists a constant $\kappa$ such that sum of the series (8.6) does not exceed $\kappa$ in value on $ab$. If $\sum_\sigma \delta_\sigma^2 < +\infty$, the absolute and uniform convergence of (8.7) on this interval is a consequence of the inequalities

$$\sum_{\sigma=N}^{N+h}\left|\frac{y_{i\sigma}(x)}{\Lambda_\sigma}\delta_\sigma\right| \leqq \left[\sum_{\sigma=N}^{N+h}\left\{\frac{y_{i\sigma}(x)}{\Lambda_\sigma}\right\}^2\right]^{1/2}\cdot\left[\sum_{\sigma=N}^{N+h}\delta_\sigma^2\right]^{1/2}$$
$$\leqq \kappa\left[\sum_{\sigma=N}^{N+h}\delta_\sigma^2\right]^{1/2}.$$

THEOREM 8.1. *For an arbitrary vector f of L the series*

$$(8.8)\qquad \phi(x) = \sum_\sigma y_\sigma(x)d_\sigma[f]$$

*converges absolutely and uniformly on ab; moreover,* $B(x)[f(x)-\phi(x)]\equiv 0$ *on this interval. For an arbitrary vector* $h(x)$ *whose components are integrable on ab,*

$$(8.9)\qquad \int_a^b hSf\,dx = \sum_\sigma \delta_\sigma[h]d_\sigma[f];$$

*in particular,*

$$(8.10)\qquad \int_a^b fS_1f\,dx = \sum_\sigma d_\sigma^2[f],$$

$$(8.11)\qquad H[f] = \sum_\sigma \Lambda_\sigma d_\sigma^2[f].$$

If $\mathcal{L}[f] = Bg$, $s[f] = 0$, it follows from (8.1) that $\Lambda_\sigma d_\sigma[f] = \delta_\sigma[g]$, $(\sigma=1, 2, \cdots)$, and the absolute and uniform convergence of (8.8) is a consequence of Lemmas 8.1 and 8.2. Clearly, $\psi(x)\equiv S(x)[f(x)-\phi(x)]$ satisfies $\delta_\sigma[\psi]=0$, $(\sigma=1, 2, \cdots)$. We will now show that also

$$(8.12)\qquad \int_a^b K_1(x, t)\psi(t)\,dt = 0.$$

Let $f^*$ be the solution of $\mathcal{L}[f^*]=B_1f$, $s[f^*]=0$. Then by Theorem 3.1 of Bliss [2] the series $\phi^*(x)=\sum_\sigma y_\sigma(x)\,d_\sigma[f^*]$ converges absolutely and uniformly on $ab$, and $B_1[f^*-\phi^*]\equiv 0$ on this interval. This latter condition, in view of the first paragraph of this section, implies $S[f^*-\phi^*]\equiv 0$. Consequently, since $\Lambda_\sigma d_\sigma[f^*]=d_\sigma[f]$, $(\sigma=1, 2, \cdots)$, we have

$$0 \equiv S(x)\left[f^*(x) - \sum_\sigma \frac{y_\sigma(x)}{\Lambda_\sigma}d_\sigma[f]\right] \equiv \int_a^b K_1(x, t)\psi(t)\,dt,$$

the latter relation being verified by direct computation. Finally, as

$$\psi(x) = \int_a^b g(t)K_1(t, x)\,dt - \sum_\sigma u_\sigma(x)d_\sigma[f],$$

it follows from (8.12) and $\delta_\sigma[\psi]=0$, $(\sigma=1, 2, \cdots)$, that $\int_a^b\psi\psi\,dx=0$, and hence $\psi\equiv 0$ on $ab$. In particular, $0\equiv\tilde{T}^{-1}S[f-\phi]\equiv B[f-\phi]$.

Equations (8.9) and (8.10) are ready consequences of the relation $Sf\equiv S\phi$ and the uniform convergence of the series $\phi$. Relation (8.11) in turn results from (8.9), the conditions $\Lambda_\sigma d_\sigma[f]=\delta_\sigma[g]$, and $H[f]=\int_a^b gSf\,dx$.

In view of (8.10) we also have the following result.

COROLLARY 1. *A vector $f$ of $L$ satisfies $\int_a^b fS_1y\,dx=0$ with every characteristic solution $y$ of* (8.1) *if and only if $Bf\equiv 0$ on $ab$.*

Corollary 2.2 of Bliss [2], when applied to the system (8.1), would imply the result of the preceding corollary for vectors $f$ belonging to $L_1$, instead of to the space $L$.

COROLLARY 2. *If the vector $f$ belongs to $L$, and $f^*$ in turn satisfies $\mathcal{L}[f^*]=B_1f$, $s[f^*]=0$, then $f^*(x)=\sum_\sigma y_\sigma(x)\,d_\sigma[f^*]$.*

In view of the uniform convergence of the series (8.8) associated with $f$, and the relation $\Lambda_\sigma d_\sigma[f^*]=d_\sigma[f]$, $(\sigma=1, 2, \cdots)$, we have

$$\begin{aligned} f^*(x) &= \int_a^b K(x, t)S(t)f(t)\,dt \\ &= \sum_\sigma \left\{\int_a^b K(x, t)S(t)y_\sigma(t)\,dt\right\} d_\sigma[f] \\ &= \sum_\sigma \frac{y_\sigma(x)}{\Lambda_\sigma} d_\sigma[f] \\ &= \sum_\sigma y_\sigma(x)d_\sigma[f^*]. \end{aligned}$$

It is to be mentioned that the above relation $Sf\equiv\sum_\sigma u_\sigma\delta_\sigma[Sf]=\sum_\sigma u_\sigma\,d_\sigma[f]$ for a vector $f$ of $L$, and the subsequent proof of (8.10), (8.11), could have been taken directly from the Hilbert-Schmidt theory of integral equations. However, by the above treatment we have proved more; namely, the absolute and uniform convergence of the series (8.8) involving the characteristic solutions of the considered boundary value problem (8.1).

If the characteristic values of (8.1) are bounded below, for this particular problem the result of Theorem 7.2 may be strengthened in that the space $L$ for the problem (2.1) may essentially be substituted for the space $L_1$ belonging to (8.1). For suppose that the characteristic values are bounded below and that $\{y_\sigma, \Lambda_\sigma\}$ are so ordered that $\Lambda_1\leqq\Lambda_2\leqq\cdots$. If $\Gamma_1$ denote the totality of vectors $f$ of $L$ satisfying $\int_a^b fS_1f\,dx=1$ and $\Gamma_1$ is nonvacuous, it follows from (8.10), (8.11) that $\Lambda_1$ is the minimum of $H[f]$ in $\Gamma_1$; moreover, in view of Corollary 1, it follows by an argument similar to that used in the proof of Theorem 7.2 that this minimum is attained by a particular $f$ if and only if $f=Y_1+\Phi_1$, where $Y_1$ is a characteristic solution of (8.1) for $\lambda=\Lambda_1$ and $\Phi_1$ is a vector of $L$ such that $B\Phi_1\equiv 0$. In general, if $\Lambda_1, \cdots, \Lambda_{m-1}$ exist for (8.1), let $\Gamma_m$ de-

note the totality of vectors $f$ of $L$ satisfying $\int_a^b fS_1f\,dx=1$, $d_\sigma[f]=0$, $(\sigma=1,\cdots,m-1)$. If this class is nonvacuous, we have by a corresponding argument that $\Lambda_m$ exists and is the minimum of $H[f]$ in $\Gamma_m$; moreover, this minimum is attained by a particular $f$ of $\Gamma_m$ if and only if $f$ is of the form $Y_m+\Phi_m$, where $Y_m$ is a characteristic solution for $\lambda=\Lambda_m$ and $\Phi_m$ is a vector of $L$ satisfying $B\Phi_m\equiv 0$ on $ab$.

From the above Corollary 1 and equation (8.11) we deduce the following theorem.

THEOREM 8.2. *A system* (2.1) *which satisfies conditions* (i), (ii) *and* (iv) *of* §2 *also satisfies condition* (iii)′, *and is consequently H-definitely self-adjoint, if and only if the corresponding system* (8.1) *has no characteristic values* $\Lambda$ *satisfying* $\Lambda\leqq 0$.

It is to be emphasized that there is no assurance under the conditions of this theorem that the system (8.1) shall have any characteristic values. For example, the system

$$y_1' = 0, \qquad y_2' = -\lambda y_1, \qquad y_1(0) = 0 = y_2(1)$$

is not only definitely self-adjoint (Bliss [2, p. 427]), but also $H$-definitely self-adjoint with the matrix (5.13), whereas this system has no characteristic values. Moreover, the corresponding system (8.1) is identical with the given system and thus possesses no characteristic values.

If an $H$-definitely self-adjoint system has $k$ linearly independent characteristic solutions it is a consequence of Theorem 6.1 and formula (8.11) that the corresponding system (8.1) has at least $k$ linearly independent characteristic solutions. In general, however, when (2.1) is $H$-definitely self-adjoint the corresponding system (8.1) may have more linearly independent characteristic solutions than the original system. To illustrate this possibility, consider the example (5.11) where, as in §5, it is supposed that $\int_0^1 b(t)\,dt=0$. The corresponding system (8.1) is

$$y_1' = 0, \qquad y_2' = -\Lambda b^2(x)y_1,$$
$$y_1(0) - y_2(0) = 0, \qquad y_1(1) + y_2(1) = 0,$$

and this system is seen to have the single characteristic value $\Lambda=2/\int_0^1 b^2(t)\,dt$ of index one, whereas the original system (5.11) has no characteristic values.

THEOREM 8.3. *For an H-definitely self-adjoint system* (2.1) *there exists a constant* $d>0$ *such that the inequality*

$$\int_a^b fS_1f\,dx \leqq dH[f] \tag{8.13}$$

*holds for arbitrary vectors f of L.*

If the corresponding system (8.1) admits of characteristic values, then in view of Theorem 8.2 and the minimizing properties of the characteristic values, inequality (8.13) holds for $d$ the reciprocal of the smallest characteristic value. If (8.1) possesses no characteristic values, then $Bf \equiv 0$ on $ab$ for every vector $f$ of $L$, the two integrals appearing in (8.13) are individually zero, and in this case $d$ may be chosen as an arbitrary positive number.

Since for an $H$-definitely self-adjoint system (2.1) the elements of the matrix $K_1(x, t)$ are continuous on $a \leqq x, t \leqq b$, and the quadratic functional (4.3) is positive semi-definite for arbitrary vectors $g$, we have the following theorem of Mercer (Mercer [8]; also, for example, [3, p. 456]).

THEOREM 8.4. *If* (2.1) *is $H$-definitely self-adjoint the series*

$$\sum_{\sigma} \frac{u_{i\sigma}(x)u_{j\sigma}(t)}{\Lambda_\sigma}, \qquad i, j = 1, 2, \cdots, n, \tag{8.14}$$

*converges absolutely and uniformly on $a \leqq x, t \leqq b$ and has the sum $K_{1ij}(x, t)$.*

9. **Further results for $H$-definitely self-adjoint systems.** The conclusions of the previous section will now be used in the proof of additional results for an $H$-definitely self-adjoint problem (2.1). For such a problem let $\mathbf{C}_1$ denote the totality of vectors $f$ of $L$ satisfying the condition

$$\int_a^b fSf\,dx = 1. \tag{9.1}$$

THEOREM 9.1. *If for an $H$-definitely self-adjoint problem* (2.1) *the class $\mathbf{C}_1$ is nonvacuous, then this system possesses positive characteristic values; moreover, the smallest positive characteristic value is the minimum of $H[f]$ in the class $\mathbf{C}_1$.*

If the class $\mathbf{C}_1$ is nonvacuous, let $\lambda_1$ denote the greatest lower bound of $H[f]$ in this class; it then follows that

$$H[f:\lambda_1] \equiv H[f] - \lambda_1 \int_a^b fSf\,dx \geqq 0$$

for arbitrary vectors $f$ of $L$. In view of relation (2.5) for a characteristic solution, there is clearly no positive characteristic value of (2.1) less than $\lambda_1$. Hence we have only to prove that $\lambda_1$ is a characteristic value.

Theorem 9.1 will be established by indirect argument. Let $f_m \equiv (f_{im})$, $(m = 1, 2, \cdots)$, be a sequence of vectors of $\mathbf{C}_1$ such that $\lim_{m\to\infty} H[f_m] = \lambda_1$; on the assumption that $\mathbf{C}_1$ is not empty such a sequence exists. Now suppose that $\lambda = \lambda_1$ is not a characteristic value; then there exist unique corresponding vectors $h_m(x)$ such that

$$\mathcal{L}[h_m] - \lambda_1 B h_m = Bf_m, \qquad s[h_m] = 0, \qquad m = 1, 2, \cdots. \tag{9.2}$$

Moreover, if $G(x, t:\lambda_1)$ denote the Green's matrix for the incompatible system $\mathcal{L}[y]-\lambda_1 By=0$, $s[y]=0$, we have

$$h_m(x) = \int_a^b H(x, t:\lambda_1)S(t)f_m(t)\,dt, \qquad m = 1, 2, \cdots,$$

where $H(x, t:\lambda_1)=G(x, t:\lambda_1)\tilde{T}^{-1}(t)$. By an elementary vector inequality,

$$\text{norm }\{h_m(x)\} \leqq \kappa\int_a^b \text{norm }\{S(t)f_m(t)\}\,dt = \kappa\int_a^b [f_m S_1 f_m]^{1/2}dt,$$

for $\kappa$ a suitable constant depending only upon the bounds of the elements of $H(x, t:\lambda_1)$ on $a \leqq x, t \leqq b$. By the use of Schwarz' inequality and Theorem 8.3 it then follows that

$$\text{norm }\{h_m(x)\} \leqq \kappa\{(b-a)dH[f_m]\}^{1/2}. \tag{9.3}$$

In particular, since $\{f_m\}$ is a minimizing sequence of $\boldsymbol{C}_1$, the sequence $\{H[f_m]\}$ is uniformly bounded and there exists a constant $\kappa_1$ such that

$$\int_a^b [\text{norm }\{h_m(x)\}]^2\,dx \leqq \kappa_1, \qquad m = 1, 2, \cdots. \tag{9.4}$$

Now set $g_m(x)=f_m(x)+c\,h_m(x)$, $(m=1, 2, \cdots)$, where $c$ is a real constant. Then $g_m$ is a vector of $L$, and

$$\begin{aligned} H[g_m:\lambda_1] &= H[f_m:\lambda_1] + 2cH[f_m;\,h_m:\lambda_1] + c^2H[h_m:\lambda_1] \\ &= H[f_m:\lambda_1] + 2c + c^2\int_a^b h_m S f_m\,dx, \end{aligned}$$

in view of (9.2) and the fact that $f_m$ belongs to the class $\boldsymbol{C}_1$. Now

$$\begin{aligned} \left|\int_a^b h_m S f_m\,dx\right| &\leqq \int_a^b [\text{norm }\{h_m\}]\cdot[\text{norm }\{Sf_m\}]\,dx \\ &\leqq \frac{1}{2}\int_a^b [h_m h_m + f_m S_1 f_m]\,dx \\ &\leqq \frac{1}{2}(\kappa_1 + dH[f_m]), \end{aligned}$$

by (9.4) and Theorem 8.3. Consequently, since $\{H[f_m]\}$ is uniformly bounded, there exists a constant $\kappa_2$ such that

$$\left|\int_a^b h_m S f_m\,dx\right| \leqq \kappa_2, \qquad m = 1, 2, \cdots.$$

Now let $c$ be a value such that $2c+c^2\kappa_2<0$; that is, $0>c>-2/\kappa_2$. As

$\lim_{m\to\infty} H[f_m:\lambda_1]=0$, for $m$ sufficiently large it follows that $H[g_m:\lambda_1]<0$, contrary to the definition of $\lambda_1$. Hence $\lambda_1$ is a characteristic value and the theorem is proved.

Let $C_{-1}$ denote the totality of vectors $f$ of $L$ satisfying

$$\int_a^b fSf\,dx = -1.$$

If the matrix $B(x)$ is replaced by $-B(x)$, then the class $C_{-1}$ for the original problem corresponds to the class $C_1$ for the modified problem; clearly such a substitution does not affect the $H$-definite self-adjointness of the problem. Hence we have the following result.

COROLLARY. *If for an H-definitely self-adjoint problem the class $C_{-1}$ is nonvacuous, then this system possesses negative characteristic values; moreover, if $\lambda_{-1}$ denote the largest negative characteristic value, then $-\lambda_{-1}$ is the minimum of $H[f]$ in the class $C_{-1}$.*

For convenience, in the remainder of this section we shall denote the totality of positive characteristic values of (2.1) by $\{\lambda_m\}$, $(m=1, 2, \cdots)$, each repeated a number of times equal to its multiplicity and ordered so that $\lambda_1 \leqq \lambda_2 \leqq \cdots$. Similarly, $\{\lambda_{-m}\}$, $(m=1, 2, \cdots)$, denotes the totality of negative characteristic values, each repeated a number of times equal to its multiplicity and the set ordered so that $\lambda_{-1} \geqq \lambda_{-2} \geqq \cdots$. Corresponding to $\lambda_m$, $\lambda_{-m}$ we shall associate characteristic solutions $y_m$, $y_{-m}$, respectively, such that $\{y_m, y_{-m}\}$, $(m=1, 2, \cdots)$, is a maximal set of linearly independent solutions orthonormal in the sense of (6.1). Clearly either one, or both, of the sequences $\{y_m, \lambda_m\}$, $\{y_{-m}, \lambda_{-m}\}$ may be vacuous or consist of only a finite number of characteristic values and associated characteristic solutions.

Using the above notation, if $\lambda_1, \cdots, \lambda_{s-1}$ exist we shall denote by $C_s$ the totality of vectors $f$ of $L$ such that

$$\int_a^b fSf\,dx = 1, \qquad \int_a^b y_m Sf\,dx = 0, \qquad m = 1, 2, \cdots, s-1.$$

Similarly, if $\lambda_{-1}, \cdots, \lambda_{-(s-1)}$ exist the class $C_{-s}$ is defined as the totality of vectors $f$ of $L$ satisfying

$$\int_a^b fSf\,dx = -1, \qquad \int_a^b y_{-m} Sf\,dx = 0, \qquad m = 1, 2, \cdots, s-1.$$

THEOREM 9.2. *If for an H-definitely self-adjoint system the class $C_s$ $[C_{-s}]$ is nonvacuous, then the characteristic value $\lambda_s$ $[\lambda_{-s}]$ exists; moreover, $\lambda_s$ $[-\lambda_{-s}]$ is the minimum of $H[f]$ in the class $C_s$ $[C_{-s}]$.*

In view of the artifice used in deducing the above corollary, it suffices to restrict our attention to the case of positive characteristic values. The result

of the theorem might be established by an argument similar to that utilized by the author [7] in proving a corresponding result for special boundary value problems associated with the calculus of variations. However, the following method, which has also been used in considering accessory boundary problems of the calculus of variations, seems more elegant.

Consider the auxiliary boundary problem involving $n+2(s-1)$ variables $(y, u, v)\equiv(y_i, u_\alpha, v_\alpha)$, $(i=1, 2, \cdots, n;\ \alpha=1, 2, \cdots, s-1)$, and consisting of the differential equations and boundary conditions

$$
\begin{aligned}
y_i' &= A_{ij}(x)y_j + B_{ij}(x)y_{j\beta}(x)u_\beta + \lambda B_{ij}(x)y_j,\\
u_\alpha' &= 0,\\
v_\alpha' &= y_{i\alpha}(x)S_{ij}(x)y_j,\\
M_{ij}y_j(a) &+ N_{ij}y_j(b) = 0,\\
&v_\alpha(a) = 0,\\
&v_\alpha(b) = 0,
\end{aligned}
\tag{9.5}
$$

where the indices $\alpha, \beta$ range from 1 to $s-1$. If capital German letters denote the matrices of (9.5) corresponding to $A$, $B$, $M$ and $N$ of the system (2.1), we have

$$
\mathfrak{A} = \left\| \begin{array}{ccc} A_{ij} & B_{ij}y_{j\beta} & 0_{i\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & 0_{\alpha\beta} \\ y_{i\alpha}S_{ij} & 0_{\alpha\beta} & 0_{\alpha\beta} \end{array} \right\|, \qquad
\mathfrak{B} = \left\| \begin{array}{ccc} B_{ij} & 0_{i\beta} & 0_{i\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & 0_{\alpha\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & 0_{\alpha\beta} \end{array} \right\|,
$$

$$
\mathfrak{M} = \left\| \begin{array}{ccc} M_{ij} & 0_{i\beta} & 0_{i\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & \delta_{\alpha\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & 0_{\alpha\beta} \end{array} \right\|, \qquad
\mathfrak{N} = \left\| \begin{array}{ccc} N_{ij} & 0_{i\beta} & 0_{i\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & 0_{\alpha\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & \delta_{\alpha\beta} \end{array} \right\|.
$$

This system is seen to be self-adjoint with the matrix

$$
\mathfrak{T} = \left\| \begin{array}{ccc} T_{ij} & 0_{i\beta} & 0_{i\beta} \\ 0_{\alpha j} & 0_{\alpha\beta} & -\delta_{\alpha\beta} \\ 0_{\alpha j} & \delta_{\alpha\beta} & 0_{\alpha\beta} \end{array} \right\|,
$$

where $T\equiv\|T_{ij}\|$ is the matrix with which (2.1) is $H$-definitely self-adjoint. Condition (ii) of §2 is seen to be satisfied by this system. Since $\mathcal{L}[y_\alpha]=\lambda_\alpha By_\alpha$, $s[y_\alpha]=0$, $(\alpha=1, \cdots, s-1)$, it also follows readily that if $(y, u, v)$ is a characteristic solution of (9.5), then $u_\alpha=0$, $(\alpha=1, \cdots, s-1)$; moreover, for such a characteristic solution $y\not\equiv 0$ on $ab$. In particular, for $\lambda=0$ this result implies that whenever condition (iv) is satisfied by (2.1) this condition also holds for (9.5). Finally, if $(y, u, v)$ belongs to the corresponding linear vector space $\mathfrak{L}$ for (9.5), then $y$ belongs to the space $L$ for (2.1); also, for such $(y, u, v)$ the corresponding functional $H[y, u, v]$ reduces simply to $H[y]$. Therefore, con-

dition (iii)′ for (2.1) implies the corresponding condition for (9.5), and if (2.1) is $H$-definitely self-adjoint so also is the latter system.

Now if $f$ belongs to the class $C_s$ for (2.1), the set $\{f_i, u_\alpha \equiv \text{constant}, v_\alpha = \int_a^x y_\alpha(t)S(t)f(t)\,dt\}$ belongs to the corresponding class $\mathfrak{C}_1$ for (9.5); conversely, if $(f_i, u_\alpha, v_\alpha)$ belongs to $\mathfrak{C}_1$ the vector $f$ belongs to $C_s$. In particular, $C_s$ is vacuous if and only if $\mathfrak{C}_1$ is vacuous. If $\mathfrak{C}_1$ is nonvacuous, then by Theorem 9.1 the minimum of $H[y, u, v] \equiv H[y]$ in this class exists and is the smallest positive characteristic value of (9.5). Since, as pointed out above, for a characteristic solution of (9.5) we have $u \equiv 0$, $y \not\equiv 0$, it follows that the smallest positive characteristic value of (9.5) is a characteristic value for (2.1). It is obvious that the characteristic value thus determined is equal to $\lambda_s$ according to the previously introduced notation.

We are now in a position to derive a result which is complementary to that of Theorem 6.1.

THEOREM 9.3. *A necessary and sufficient condition that an $H$-definitely self-adjoint system have at least $k$ positive [negative] characteristic values, where it is to be understood that each such value is counted a number of times equal to its multiplicity, is that the quadratic functional $\int_a^b fSf\,dx$ be positive [negative] definite on a linear subspace of $L$ of dimension $k$.*

For suppose that positive characteristic values $\lambda_1, \cdots, \lambda_k$ exist for such a system (2.1), and that $y_1, \cdots, y_k$ are corresponding orthonormal characteristic solutions. Then for arbitrary constants $(d_1, \cdots, d_k) \neq (0, \cdots, 0)$ we have for $f = y_1d_1 + \cdots + y_kd_k$ that $\int_a^b fSf\,dx = d_1^2 + \cdots + d_k^2 > 0$. On the other hand, if $\int_a^b fSf\,dx$ is positive definite on a linear subspace of $L$ of dimension $k$, then the classes $C_1, \cdots, C_k$ as defined above are seen to be nonvacuous and (2.1) has at least $k$ positive characteristic values. Again, in view of the possibility of replacing $B$ by $-B$, the result for negative characteristic values is a ready consequence of the result for positive characteristic values.

We shall now give a particular condition which is sufficient to insure that an $H$-definitely self-adjoint system (2.1) has infinitely many characteristic values of a given sign. We shall denote by $(v_+)$ the following hypothesis.

$(v_+)$ There is a subinterval $a_1b_1$, $a < a_1 < b_1 < b$, of $ab$ such that if $a_1'$, $b_1'$ are arbitrary values satisfying $a_1 \leqq a_1' < b_1' \leqq b_1$, then there exists a vector $g(x)$ and associated $f(x)$ satisfying $\mathfrak{L}[f] = Bg$ for which $f \equiv 0$ outside $a_1'b_1'$, whereas $\int_a^b fSf\,dx > 0$.

The condition obtained by replacing in $(v_+)$ the relation "$\int_a^b fSf\,dx > 0$" by "$\int_a^b fSf\,dx < 0$" will be referred to as $(v_-)$. Using the device of the proof of Theorem 6.2, and the result of the above theorem, one obtains the following conclusion.

THEOREM 9.4. *If an $H$-definitely self-adjoint system* (2.1) *satisfies hypothesis* $(v_+)$, *then this system admits infinitely many positive characteristic values. Simi-*

*larly, if such a system satisfies condition* ($v_-$), *there exist infinitely many negative characteristic values.*

In agreement with our modified notation for the characteristic values and solutions of an $H$-definitely self-adjoint system (2.1), we write

$$c_\mu[f] = \frac{|\lambda_\mu|}{\lambda_\mu}\int_a^b fSy_\mu\,dx, \quad \mu = 1, -1, 2, -2, \cdots.$$

THEOREM 9.5. *For an arbitrary vector $f$ of $L$,*

$$\int_a^b fSf\,dx = \sum_{\mu=-\infty}^{\infty} \frac{|\lambda_\mu|}{\lambda_\mu} c_\mu^2[f]; \tag{9.6}$$

*moreover, if $f$ and $h$ are vectors of $L$,*

$$\int_a^b fSh\,dx = \sum_{\mu=-\infty}^{\infty} \frac{|\lambda_\mu|}{\lambda_\mu} c_\mu[f]c_\mu[h]. \tag{9.7}$$

In view of Theorem 9.2, relation (9.6) is readily seen to be true if (2.1) admits only a finite number of characteristic values. We shall prove this relation for the case in which this system has infinitely many positive, and also infinitely many negative, characteristic values; the modification in the proof whenever the system has only a finite number of characteristic values of one sign is obvious.

Corresponding to a vector $f$ of $L$ and a given positive integer $k$, set $f_k = f - \sum_{\mu=-k}^{k} y_\mu(x)\, c_\mu[f]$. Then $c_\mu[f_k]=0$, ($\mu = -k, \cdots, k$), and by the minimizing properties of the characteristic values of (2.1) we have

$$H[f_k] \geqq \lambda_{k+1}\int_a^b f_k S f_k\,dx \geqq 0$$

if $\int_a^b f_k S f_k\,dx \geqq 0$; whereas

$$H[f_k] \geqq \lambda_{-k-1}\int_a^b f_k S f_k\,dx \geqq 0$$

if $\int_a^b f_k S f_k\,dx \leqq 0$. Consequently,

$$\left|\int_a^b f_k S f_k\,dx\right| \leqq \max\left\{\frac{H[f_k]}{\lambda_{k+1}},\ -\frac{H[f_k]}{\lambda_{-k-1}}\right\}.$$

Moreover, since

$$0 \leqq H[f_k] = H[f] - \sum_{\mu=-k}^{k} |\lambda_\mu|\, c_\mu^2[f] \leqq H[f],$$

it follows that

$$0 = \lim_{k\to\infty} \int_a^b f_k S f_k \, dx = \int_a^b fSf \, dx - \lim_{k\to\infty} \sum_{\mu=-k}^{k} \frac{|\lambda_\mu|}{\lambda_\mu} c_\mu^2[f].$$

Since the series involved in (9.6) clearly converges absolutely, this relation is established. Relation (9.7) is then immediate since if $f$ and $h$ are vectors of $L$, so also are $f+h$ and $f-h$, and

$$\int_a^b fSh \, dx = \frac{1}{4}\left[\int_a^b (f+h)S(f+h)dx - \int_a^b (f-h)S(f-h)dx\right],$$
$$c_\mu[f \pm h] = c_\mu[f] \pm c_\mu[h].$$

COROLLARY 1. *If $f$ is a vector of $L^2$ then the equality sign holds in* (6.2), *that is*

$$H[f] = \sum_{\mu=-\infty}^{\infty} |\lambda_\mu| c_\mu^2[f]. \tag{9.8}$$

Let $h$ be a vector of $L$ such that $\mathcal{L}[f] = Bh$, $s[f]=0$. Since $\lambda_\mu c_\mu[f] = c_\mu[h]$, $(\mu = 1, -1, 2, -2, \cdots)$, this corollary is a consequence of (9.7) and the relation $H[f] = \int_a^b fSh \, dx$.

It is to be noted that in general we do not have relation (9.8) for arbitrary vectors $f$ of the space $L$. This fact is shown by the example (5.11), in view of the comment immediately preceding Theorem 8.3. We do have, however, the following result.

COROLLARY 2. *If the class $C_s$, $(s=1, -1, 2, -2, \cdots)$, is nonvacuous for an $H$-definitely self-adjoint system* (2.1), *then the minimum of $H[f]$ in this class is attained by a particular $f$ of $C_s$ if and only if $f = Y_s(x) + \Phi_s(x)$, where $Y_s(x)$ is a characteristic solution for $\lambda = \lambda_s$ and $\Phi_s$ is a vector of $L$ satisfying $B\Phi_s \equiv 0$ on $ab$.*

By the use of inequality (6.2), relation (9.6), and an argument similar to that employed in the proof of Theorem 7.2, we have that if $f$ is a vector of a nonvacuous class $C_s$ which renders $H[f]$ its minimum value in this class, then $f = Y_s + \Phi_s$, where $Y_s$ is a characteristic solution of (2.1) for $\lambda = \lambda_s$ and $\Phi_s$ is a vector of $L$ such that $c_\mu[\Phi_s] = 0$, $(\mu = 1, -1, 2, -2, \cdots)$. It then follows from Corollary 1 to Theorem 5.4 that $\int_a^b \Phi_s Sy \, dx = 0$ for every vector $y$ of $L$ and thus, in particular, $\int_a^b \Phi_s S\Phi_s \, dx = 0$. Then

$$\begin{aligned} 0 = H[f; \lambda_s] &= H[Y_s : \lambda_s] + 2H[\Phi_s; Y_s : \lambda_s] + H[\Phi_s : \lambda_s] \\ &= H[\Phi_s : \lambda_s] = H[\Phi_s], \end{aligned}$$

and hence $B(x)\Phi_s(x) \equiv 0$ on $ab$ by condition (iii)'.

THEOREM 9.6. *If* (2.1) *is $H$-definitely self-adjoint and we write $v_\mu(x) = S(x)y_\mu(x)$, $(\mu = 1, -1, 2, -2, \cdots)$, then each of the series*

$$\sum_{\mu=-\infty}^{\infty} \frac{[v_{i\mu}(x)]^2}{|\lambda_\mu|}, \qquad i = 1, 2, \cdots, n, \tag{9.9}$$

*converges and its sum does not exceed $K_{1ii}(x, x)$ on ab. For fixed values of one of the variables x, t, each of the series*

$$\sum_{\mu=-\infty}^{\infty} \frac{v_{i\mu}(x)v_{j\mu}(t)}{|\lambda_\mu|}, \qquad i, j = 1, 2, \cdots, n, \tag{9.10}$$

*converges absolutely and uniformly in the other variable on ab. Finally, for an arbitrary vector f of L the vector series*

$$\sum_{\mu=-\infty}^{\infty} v_\mu(x)c_\mu[f] \tag{9.11}$$

*converges absolutely and uniformly on ab.*

Corresponding to an arbitrary vector $g(x)$, define $f$ by $\mathcal{L}[f]=Bg$, $s[f]=0$. Then for an arbitrary integer $k$,

$$\begin{aligned} 0 &\leqq H\left[f - \sum_{|\mu|\leqq k} y_\mu c_\mu[f]\right] \\ &\equiv \int_a^b \int_a^b g_i(x)\left\{K_{1ij}(x, t) - \sum_{|\mu|\leqq k} \frac{v_{i\mu}(x)v_{j\mu}(t)}{|\lambda_\mu|}\right\} g_j(t)\, dx\, dt. \end{aligned} \tag{9.12}$$

Applying the argument of Lemma 4.2 to the double integral of (9.12), it follows in particular that

$$\sum_{|\mu|\leqq k} \frac{[v_{i\mu}(x)]^2}{|\lambda_\mu|} \leqq K_{1ii}(x, x)$$

for each integer $k$. Hence the series (9.9) converges and its sum does not exceed $K_{1ii}(x, x)$. Since the sum of this series is uniformly bounded on $ab$, Cauchy's inequality insures that each of the series (9.10) converges absolutely and uniformly in each of the variables on $ab$ for fixed values of the other variable. In particular, each of these series defines a function which is continuous in each of the variables $x$, $t$ separately on $ab$. Since for an arbitrary $f$ of $L$ the series $\sum_\mu |\lambda_\mu|\ c_\mu^2[f]$ converges, the absolute and uniform convergence of (9.11) is a consequence of the uniform boundedness of the sum of the series (9.9) on $ab$ and Cauchy's inequality.

The proof of the above convergence properties of (9.9), (9.10) parallels that of corresponding results (see, for example, [3, p. 456]) used in establishing Theorem 8.4 for the boundary problem (8.1). We are unable to prove for (2.1) a result as general in character as Theorem 8.4 gives for system (8.1), however, since for an $H$-definite self-adjoint system we do not in general

have that relation (9.8) is valid for arbitrary vectors $f$ of $L$. When this latter condition is fulfilled for a particular $H$-definite self-adjoint problem it then follows that the sum of the series (9.10) is $K_{1ij}(x, t)$; in particular, the sum of (9.9) is $K_{1ii}(x, x)$, by Dini's theorem the convergence of this latter series is uniform on $ab$, and it then follows that the series (9.10) converges absolutely and uniformly in $(x, t)$ jointly.

THEOREM 9.7. *For a vector $f$ of $L$ relation* (9.8) *holds if and only if*

$$S(x)f(x) = \sum_{\mu=-\infty}^{\infty} S(x)y_\mu(x)c_\mu[f] \equiv \sum_{\mu=-\infty}^{\infty} v_\mu(x)c_\mu[f]. \tag{9.13}$$

From the above theorem we know that the right-hand member of (9.13) converges absolutely and uniformly on $ab$, and hence defines a vector whose components are continuous on this interval. If $\mathcal{L}[f]=Bg$, $s[f]=0$, and relation (9.13) holds for $f$, then

$$H[f] = \int_a^b gSf\,dx = \sum_{\mu=-\infty}^{\infty} c_\mu[f] \int_a^b gSy_\mu\,dx = \sum_{\mu=-\infty}^{\infty} |\lambda_\mu|\, c_\mu^2[f],$$

since $c_\mu[g] = \lambda_\mu c_\mu[f]$. On the other hand, if we define $f_k(x) = f(x) - \sum_{|\mu|\leq k} y_\mu(x)\, c_\mu[f]$, relation (9.8) is equivalent to the condition $\lim_{k\to\infty} H[f_k] = 0$, whereas (9.13) is equivalent to $F(x) \equiv \lim_{k\to\infty} S(x)f_k(x) \equiv 0$. In view of relation (8.13) for $f_k$ it then follows that if (9.8) holds for a vector $f$ then the associated vector $F$ is identically zero, that is, relation (9.13) is also valid.

Since the matrix $T$ is nonsingular, it is clear that relation (9.13) holds for a particular $f$ if and only if the series $\sum_\mu B(x)y_\mu(x)c_\mu[f]$ converges absolutely and uniformly on $ab$, and

$$B(x)f(x) = \sum_{\mu=-\infty}^{\infty} B(x)y_\mu(x)c_\mu[f]. \tag{9.14}$$

COROLLARY. *If $f$ is a vector of $L^2$, then the series $\phi(x)=\sum_\mu y_\mu(x)\, c_\mu[f]$ converges absolutely and uniformly on $ab$, and $B(x)[f(x)-\phi(x)]\equiv 0$.*

Corollary 1 to Theorem 9.5 and Theorem 9.7 implies that relation (9.13), and hence (9.14), holds for such an $f$. We have, therefore, only to prove the absolute and uniform convergence of the series $\phi$. Let $h(x)$ be a vector of $L$ such that $\mathcal{L}[f]=Bh$, $s[f]=0$. Now since $h$ belongs to $L$, the series $\sum_\mu B(x)y_\mu(x)c_\mu[h] \equiv \sum_\mu B(x)y_\mu(x)\lambda_\mu c_\mu[f]$ converges absolutely and uniformly on $ab$. Hence the corresponding convergence of $\phi$ is a consequence of

$$\phi(x) = \sum_{\mu=-\infty}^{\infty} \left\{ \int_a^b G(x, t)B(t)y_\mu(t)\,dt \right\} \lambda_\mu c_\mu[f]$$
$$= \int_a^b G(x, t) \left\{ \sum_{\mu=-\infty}^{\infty} B(t)y_\mu(t)c_\mu[h] \right\} dt.$$

In conclusion, we shall prove the following general expansion theorem.

THEOREM 9.8. *If for an arbitrary vector f the series* $\sum_\mu B(x)y_\mu(x)\ c_\mu[f]$ *converges uniformly and satisfies relation* (9.14) *on ab, then for* $f^*(x)$ *defined by the system* $\mathcal{L}[f^*]=Bf$, $s[f^*]=0$, *the series* $\sum_\mu y_\mu(x)\ c_\mu[f^*]$ *converges uniformly on this interval to* $f^*(x)$.

This result is an immediate consequence of the relations

$$\begin{aligned} f^*(x) &= \int_a^b G(x, t)B(t)f(t)\,dt \\ &= \sum_{\mu=-\infty}^{\infty}\left\{\int_a^b G(x, t)B(t)y_\mu(t)\,dt\right\}c_\mu[f] \\ &= \sum_{\mu=-\infty}^{\infty}\frac{y_\mu(x)}{\lambda_\mu}c_\mu[f] = \sum_{\mu=-\infty}^{\infty} y_\mu(x)c_\mu[f^*]. \end{aligned}$$

10. **A boundary problem of the calculus of variations.** We shall now consider a system of the form (2.1) associated with the problem of Bolza in the calculus of variations. The symbols $\eta\equiv(\eta_i)$, $\eta'\equiv(\eta_i')$ will denote the functions $[\eta_1(x), \cdots, \eta_n(x)]$ and the set of their derivatives, respectively. Let

$$J[\eta] = 2Q[\eta(a), \eta(b)] + \int_a^b 2\omega(x, \eta, \eta')\,dx, \tag{10.1}$$

where $\omega$ and $Q$ are quadratic forms in the $2n$ variables $\eta_i$, $\eta_i'$ and $\eta_i(a)$, $\eta_i(b)$, respectively. The functional $J[\eta]$ is of the form of the second variation of a problem of Bolza. The boundary value problem to be considered consists of the Euler-Lagrange differential equations and transversality conditions for the problem of minimizing $J[\eta]$ in a class of arcs $\eta=\eta(x)$ which satisfy a set of ordinary linear differential equations of the first order

$$\Phi_\alpha(x, \eta, \eta') \equiv \Phi_{\alpha\pi_i}(x)\eta_i' + \Phi_{\alpha\eta_i}(x)\eta_i = 0, \qquad \alpha = 1, \cdots, m < n, \tag{10.2}$$

the linear homogeneous end conditions

$$\Psi_\gamma[\eta(a), \eta(b)] \equiv \Psi_{\gamma;ja}\eta_j(a) + \Psi_{\gamma;jb}\eta_j(b) = 0, \quad \gamma = 1, \cdots, p \leqq 2n, \tag{10.3}$$

and which render a fixed constant value to the integral

$$\int_a^b \eta_i\mathfrak{K}_{ij}(x)\eta_j\,dx. \tag{10.4}$$

For the general problem of Bolza the second variation may be written as (10.1) if one includes in the set $\eta$ not only the variations of the dependent functions in the original problem of Bolza, but also two additional functions representing the variations of the end values; these latter two functions are further restricted by including in (10.2) two additional differential equations which require them to be constant on $ab$.

Throughout the present section the following subscripts have the ranges indicated: $i, j, k=1, \cdots, n$; $\alpha, \beta=1, \cdots, m$; $\sigma, \tau=1, \cdots, 2n$; $\gamma, \nu=1, \cdots, p$; $\theta, \phi=1, \cdots, 2n-p$. Partial derivatives of $\omega(x, \eta, \pi)$, $\Phi_\alpha(x, \eta, \pi)$ with respect to the variables $\eta_i$, $\pi_i$ will be denoted by writing these variables as subscripts; correspondingly, derivatives of $Q$ and $\Psi_\gamma$ with respect to the arguments $\eta_i(a)$, $\eta_i(b)$ will be denoted by $Q_{ia}$, $\Psi_{\gamma;ia}$, $Q_{ib}$, $\Psi_{\gamma;ib}$, respectively.

The analysis of this section is based on the following hypotheses.

($H_1$) The coefficients of the quadratic form $\omega(x, \eta, \pi)$ and the linear forms $\Phi_\alpha(x, \eta, \pi)$ are real single-valued functions of $x$ on $ab$. The functions $\omega_{\pi_i\pi_j}$, $\omega_{\pi_i\eta_j}$, $\Phi_{\alpha\pi_j}$ are of class $C^1$, while the functions $\pi_{\eta_i\eta_j}$, $\Phi_{\alpha\eta_j}$, $\mathfrak{K}_{ij}=\mathfrak{K}_{ji}$ are continuous on this interval. Finally, the matrix $\|\Phi_{\alpha\pi_j}(x)\|$ is of rank $m$ on $ab$, the coefficients of the quadratic form $Q$ and the linear forms $\Psi_\gamma$ are real constants, and the matrix $\|\Psi_{\gamma;ja}\Psi_{\gamma;jb}\|$ has rank $p$.

($H_2$) The matrix

$$\begin{Vmatrix} \omega_{\pi_i\pi_j} & \Phi_{\beta\pi_i} \\ \Phi_{\alpha\pi_j} & 0_{\alpha\beta} \end{Vmatrix} \tag{10.5}$$

is nonsingular on $ab$.

An arc $\eta$ will be termed *differentially admissible* if its components $\eta_i(x)$ are of class $D^1$ on $ab$, and satisfy $\Phi_\alpha=0$ on this interval. An arc whose end values at $a$ and $b$ satisfy $\Psi_\gamma=0$ will be called *terminally admissible*; finally, an arc which is both differentially and terminally admissible will be said to be *admissible*.

($H_3$) There exist $p$ differentially admissible arcs $\eta_i=\eta_{i\nu}$, such that the determinant $|\Psi_\gamma[\eta_\nu(a), \eta_\nu(b)]|$ is different from zero.

For arbitrary constants $\mu_\alpha$ define

$$\Omega(x, \eta, \pi, \mu) = \omega(x, \eta, \pi) + \mu_\alpha\Phi_\alpha(x, \eta, \pi). \tag{10.6}$$

Under the hypotheses ($H_1$), ($H_2$), ($H_3$) it follows from the theory of the problem of Bolza that if $\eta(x)$ is a minimizing arc for the above defined calculus of variations problem, then there exist multipliers $\lambda=$ constant, $\mu_\alpha=\mu_\alpha(x)$ such that the set $[\eta_i(x), \mu_\alpha(x), \lambda]$ satisfies the differential equations

$$\begin{aligned} (d/dx)\Omega_{\pi_i}(x, \eta, \eta', \mu) - \Omega_{\eta_i}(x, \eta, \eta', \mu) + \lambda\mathfrak{K}_{ij}(x)\eta_j &= 0, \\ \Phi_\alpha(x, \eta, \eta') &= 0; \end{aligned} \tag{10.7}$$

moreover, there exist constants $d_\gamma$ satisfying

$$\begin{aligned} Q_{ia}[\eta] + d_\gamma\Psi_{\gamma;ia} - \Omega_{\pi_i}(x, \eta, \eta', \mu)\big|^{x=a} &= 0, \\ Q_{ib}[\eta] + d_\gamma\Psi_{\gamma;ib} + \Omega_{\pi_i}(x, \eta, \eta', \mu)\big|^{x=b} &= 0, \\ \Psi_\gamma[\eta(a), \eta(b)] &= 0. \end{aligned} \tag{10.8}$$

As (10.5) is nonsingular, the set of $m+n$ equations

$$\zeta_i = \Omega_{\pi_i}(x, \eta, \pi, \mu), \quad \Phi_\alpha(x, \eta, \pi) = 0, \quad \alpha = 1, \cdots, m;\ i = 1, \cdots, n, \tag{10.9}$$

has unique solutions

$$\pi_i = \mathfrak{A}_{ij}(x)\eta_j + \mathfrak{B}_{ij}(x)\zeta_j, \qquad \mu_\alpha = l_{\alpha j}(x)\eta_j + m_{\alpha j}(x)\zeta_j. \tag{10.10}$$

Substituting these values in $\Omega_{\pi_i}(x, \eta, \pi, \mu)$, we obtain

$$\Omega_{\eta_i}(x, \eta, \pi, \mu) = \mathfrak{C}_{ij}(x)\eta_j - \mathfrak{A}_{ji}(x)\zeta_j, \tag{10.11}$$

where in view of the above hypotheses the functions $\mathfrak{A}_{ij}$, $\mathfrak{B}_{ij}$, $\mathfrak{C}_{ij}$ are continuous on $ab$; moreover, the matrices $\|\mathfrak{B}_{ij}\|$ and $\|\mathfrak{C}_{ij}\|$ are symmetric and $\|\mathfrak{B}_{ij}\|$ is of rank $n-m$ on this interval. Consequently, the differential equations (10.7) are equivalent to the system

$$\tag{10.7'} \begin{aligned} \mathcal{L}_i[\eta, \zeta] &\equiv \eta_i' - \mathfrak{A}_{ij}(x)\eta_j - \mathfrak{B}_{ij}(x)\zeta_j = 0, \\ \mathcal{L}_{n+i}[\eta, \zeta] &\equiv \zeta_i' - \mathfrak{C}_{ij}(x)\eta_j + \mathfrak{A}_{ji}(x)\zeta_j = -\lambda\mathfrak{K}_{ij}(x)\eta_j. \end{aligned}$$

Now if $c_i = c_{i\theta}$, $d_i = d_{i\theta}$, $(\theta = 1, \cdots, 2n-p)$, are linearly independent solutions of the equations $\Psi_{\gamma;j\alpha}c_j + \Psi_{\gamma;j\beta}d_j = 0$, $(\gamma = 1, \cdots, p)$, the boundary conditions (10.8) are equivalent to the linearly independent set

$$\tag{10.8'} \begin{aligned} s_\gamma[\eta, \zeta] &\equiv \Psi_\gamma[\eta(a), \eta(b)] = 0, \\ s_{p+\theta}[\eta, \zeta] &\equiv c_{i\theta}\{Q_{ia}[\eta] - \zeta_i(a)\} + d_{i\theta}\{Q_{ib}[\eta] + \zeta_i(b)\} = 0. \end{aligned}$$

The system (10.7′), (10.8′), which is clearly of the form (2.1) in $y \equiv (\eta_i, \zeta_i)$, may be shown under the above hypotheses to satisfy conditions (i), (ii) and (iv) of §2 with the matrix

$$T \equiv \begin{Vmatrix} 0_{ij} & \delta_{ij} \\ -\delta_{ij} & 0_{ij} \end{Vmatrix}.$$

For this system we have

$$A \equiv \begin{Vmatrix} \mathfrak{A}_{ij} & \mathfrak{B}_{ij} \\ \mathfrak{C}_{ij} & -\mathfrak{A}_{ji} \end{Vmatrix}, \qquad B \equiv \begin{Vmatrix} 0_{ij} & 0_{ij} \\ -\mathfrak{K}_{ij} & 0_{ij} \end{Vmatrix}, \qquad S \equiv \begin{Vmatrix} \mathfrak{K}_{ij} & 0_{ij} \\ 0_{ij} & 0_{ij} \end{Vmatrix}.$$

We now wish to consider the condition (iii)′ for such a system. The linear vector space $L$ for this problem consists of sets $(\eta, \zeta)$ which satisfy with a corresponding $w \equiv (w_i)$ the system

$$\eta_i' = \mathfrak{A}_{ij}\eta_j + \mathfrak{B}_{ij}\zeta_j, \quad \zeta_i' = \mathfrak{C}_{ij}\eta_j - \mathfrak{A}_{ji}\zeta_j - \mathfrak{K}_{ij}w_j, \quad s_\sigma[\eta, \zeta] = 0. \tag{10.12}$$

It is readily seen from (10.8) that if $s_\sigma[\eta, \zeta] = 0$, then

$$\eta_i\zeta_i\Big|_a^b = -2Q[\eta].$$

Evaluating $H = H[\eta, \zeta]$ for such a set $(\eta, \zeta)$ we find

$$H = -\eta_i\zeta_i\Big|_a^b + \int_a^b (\eta_i'\zeta_i + \eta_i\mathfrak{C}_{ij}\eta_j - \eta_i\mathfrak{A}_{ji}\zeta_j)\,dx = J[\eta],$$

in view of (10.9) and (10.11). Consequently, (iii)′ for this system reduces to the condition that $J[\eta]>0$ *for every set* $(\eta, \zeta)$ *which satisfies with an associated vector* $w$ *the system* (10.12), *and for which* $(\mathfrak{K}_{ij}\eta_j)\not\equiv(0_i)$ *on* $ab$. If $(\eta, \zeta)$ is a solution of (10.12) for a given vector $w$, then clearly $\eta$ is an admissible arc. Hence (iii)′ is certainly satisfied if the following condition holds.

($H_4$) $J[\eta]>0$ for arbitrary nonidentically vanishing admissible arcs $\eta$.

We thus see that a system of the form treated by Reid [9] for which, using the notation of that paper, the quadratic form $G[\eta(a), \eta(b)]$ is identically zero, is $H$-definitely self-adjoint.

THEOREM 10.1. *Suppose that a problem of the above sort is H-definitely self-adjoint. If the matrix* $\mathfrak{K}(x)$ *is positive [negative] definite at a point* $x_0$ *of* $ab$, *then this system has infinitely many positive [negative] characteristic values.*

On the assumption that $\mathfrak{K}(x_0)$ is positive definite, there clearly exists a subinterval $a_1b_1$, $a<a_1<b_1<b$, throughout which $\mathfrak{K}(x)$ remains positive definite. Corresponding to an arbitrary subinterval $a_1'b_1'$ of $a_1b_1$, denote by $\zeta\equiv(\zeta_{i\rho})$, $(\rho=1, \cdots, n+1)$, a set of $n+1$ vectors whose components are of class $C^1$ on $ab$, are identically zero outside the subinterval $a_1'b_1'$, and such that the vectors $(\mathfrak{B}_{ij}\zeta_{j\rho})$, $(\rho=1, \cdots, n+1)$, are linearly independent on $a_1'b_1'$. Such vectors clearly exist since $\mathfrak{B}$ is of rank $n-m$ on $ab$; in fact, all that is necessary to insure the existence of such vectors is that $\mathfrak{B}\not\equiv 0$ on $a_1'b_1'$. Now define $\eta_\rho$ as the solution of $\eta_{i\rho}'=\mathfrak{A}_{ij}(x)\eta_{j\rho}+\mathfrak{B}_{ij}(x)\zeta_{j\rho}(x)$, $\eta_{i\rho}(a)=0$. Clearly the vectors $\eta_\rho$ are linearly independent on $ab$; moreover, there exist constants $(d_1, \cdots, d_{n+1})$ not all zero such that if we set $\eta=\eta_1d_1+\cdots+\eta_{n+1}d_{n+1}$, then $\eta_i(b)=0$. This vector satisfies with $\zeta=\zeta_1d_1+\cdots+\zeta_{n+1}d_{n+1}$ the system $\eta_i'=\mathfrak{A}_{ij}\eta_j+\mathfrak{B}_{ij}\zeta_j$ on $ab$; furthermore, as $\zeta_i\equiv 0_i$ outside $a_1'b_1'$, it follows from the conditions $\eta_i(a)=0=\eta_i(b)$ that $\eta_i\equiv 0_i$ outside this subinterval. Since $\mathfrak{K}$ is nonsingular on $a_1'b_1'$ there clearly exists a corresponding $w$ such that the differential equations of (10.12) are satisfied by $(\eta, \zeta, w)$; the boundary conditions are also satisfied by $(\eta, \zeta)$ since this set vanishes at $x=a$ and $x=b$. Consequently, since on $a_1'b_1'$ we have that $\eta\not\equiv 0$ and the matrix $\mathfrak{K}$ is positive definite, while $\eta\equiv 0$ outside this subinterval, it follows that the thus determined solution $y=(\eta, \zeta)$ of (10.12) satisfies the conditions described in hypothesis ($v_+$) of §9. Hence by Theorem 9.4 the considered system has infinitely many positive characteristic values. The corresponding result for negative characteristic values is readily deducible from the above by considering the related boundary value problem obtained by replacing the matrix $\mathfrak{K}(x)$ by $-\mathfrak{K}(x)$.

11. **A particular differential system.** Krein [7] and Kamke [6] have considered a self-adjoint boundary problem of the form

$$\mathcal{L}[u] = \lambda k(x)u, \qquad U_\sigma[u] = 0, \qquad \sigma = 1, \cdots, 2n, \tag{11.1}$$

where $\mathcal{L}[u]$ is a differential operator of the form

$$\mathcal{L}[u] = \sum_{\nu=0}^{n} [l_\nu(x)u^{(\nu)}]^{(\nu)},$$

$l_n(x) \neq 0$ and $l_\nu(x)$, $(\nu = 1, \cdots, n)$, of class $C^{(\nu)}$ on $ab$, while the $U_\sigma[u]$ are independent linear forms in the end values of $u, u', \cdots, u^{(2n-1)}$ at $x=a$ and $x=b$. Each of these authors has assumed that the functional

$$\int_a^b u\mathcal{L}[u]\,dx \tag{11.2}$$

possesses certain properties of definiteness.

Krein has supposed that (11.2) is non-negative for every function $u$ which is of class $C^{(2n)}$ on $ab$ and such that $u, u', \cdots, u^{(n-1)}$ all vanish at $a$ and $b$; moreover, that the continuous function $k(x)$ occurring in (11.1) is non-negative throughout $ab$. Kamke [6, I] has assumed that (11.2) is non-negative for every function $u$ of class $C^{(2n)}$ which satisfies the boundary conditions $U_\sigma[u]=0$; moreover, that $\lambda=0$ is not a characteristic value of (11.1). In addition, Kamke has also treated the case in which the continuous function $k(x)$ changes sign on $ab$.

It will now be shown that a system (11.1) may be written as one of the type considered in the preceding section. Now $\mathcal{L}[u]$ is the Euler expression for the integral

$$\int_a^b \{l_0(x)u^2 - l_1(x)u'^2 + \cdots + (-1)^n l_n(x)[u^{(n)}]^2\}\,dx. \tag{11.3}$$

This integral, by a device familiar in the calculus of variations, is equivalent under the substitution $\eta_1 \equiv u$ to the integral

$$\int_a^b \{l_0(x)\eta_1^2 - l_1(x)\eta_2^2 + \cdots + (-1)^{n-1}\eta_n^2 + (-1)^n\eta_n'^2\}\,dx, \tag{11.4}$$

together with the auxiliary linear differential equations

$$\Phi_\alpha \equiv \eta_\alpha' - \eta_{\alpha+1} = 0, \qquad \alpha = 1, \cdots, n-1. \tag{11.5}$$

Suppose $u$ is of class $C^{(2n)}$ and satisfies the nonhomogeneous differential equation

$$\mathcal{L}[u] + f(x) = 0. \tag{11.6}$$

If we set

$$\begin{aligned} &\eta_1 \equiv u, \quad \eta_2 \equiv u', \cdots, \eta_n \equiv u^{(n-1)}, \\ &\zeta_i \equiv (-1)^i\{[l_i(x)u^{(i)}] + [l_{i+1}(x)u^{(i+1)}]' + \cdots + [l_n(x)u^{(n)}]^{(n-i)}\}, \\ &\qquad\qquad\qquad\qquad i = 1, \cdots, n, \end{aligned} \tag{11.7}$$

it is readily seen that $(\eta_i, \zeta_i)$ satisfy the first order system

$$
(11.8)\qquad
\begin{aligned}
\eta_\alpha' &= \eta_{\alpha+1},\\
\eta_n' &= \frac{(-1)^n}{l_n(x)}\zeta_n,\\
\zeta_1' &= l_0(x)\eta_1 + f(x),\\
\zeta_{1+\alpha}' &= (-1)^\alpha l_\alpha(x)\eta_{1+\alpha} - \zeta_\alpha, \qquad \alpha = 1, \cdots, n-1.
\end{aligned}
$$

Conversely, if $(\eta_i, \zeta_i)$ is a solution of (11.8) it follows that $u \equiv \eta_1$ satisfies (11.6); moreover, $u$ and its first $2n-1$ derivatives are related to $(\eta_i, \zeta_i)$ by the equations (11.7). Hence there is complete equivalence between the single linear equation (11.6) of order $2n$ and the system (11.8) of $2n$ linear differential equations of the first order. For $f(x) \equiv 0$ this latter system is the canonical form of the Euler-Lagrange equations for the integral (11.4) subject to the auxiliary differential equations (11.5).

Now the $U_\sigma[u]$ are supposed to be $2n$ independent linear forms in the end values of $u, u', \cdots, u^{(2n-1)}$ at $x=a$ and $x=b$. In view of the assumption that $l_n(x) \neq 0$ on $ab$ it follows that they may equally well be considered as independent linear forms in the end values of the corresponding $\eta_i, \zeta_i$ at $a$ and $b$; consequently, the set $U_\sigma[u]=0$ may be written as

$$
(11.9)\qquad s_\sigma[\eta, \zeta] \equiv \overset{1}{a}_{\sigma j}\eta_j(a) - \overset{1}{b}_{\sigma j}\zeta_j(a) + \overset{2}{a}_{\sigma j}\eta_j(b) + \overset{2}{b}_{\sigma j}\zeta_j(b) = 0, \qquad \sigma = 1, \cdots, 2n.
$$

If $u$ and $u^*$ are of class $C^{(2n)}$ on $ab$, it follows from the self-adjoint character of $\mathcal{L}[u]$ that (see, for example [5, p. 123])

$$
u^*\mathcal{L}[u] - u\mathcal{L}[u^*] \equiv \frac{d}{dx} P(u; u^*),
$$

where $P(u; u^*)$ is bilinear in the sets $(u, u', \cdots, u^{(2n-1)})$ and $(u^*, u^{*\prime}, \cdots, u^{*(2n-1)})$, and is the so-called *bilinear concomitant*. In particular, if $(\eta_i, \zeta_i)$ and $(\eta_i^*, \zeta_i^*)$ are defined by (11.7) for $u$ and $u^*$, respectively, it may be readily verified that

$$
P(u; u^*) \equiv \eta_i(x)\zeta_i^*(x) - \zeta_i(x)\eta_i^*(x).
$$

The self-adjoint character of the boundary conditions implies that for arbitrary functions $u, u^*$ whose end values satisfy $U_\sigma[u]=0=U_\sigma[u^*]$, $(\sigma=1, \cdots, 2n)$, we have $P(u; u^*)\big|_a^b=0$. Consequently, if $(\eta_i, \zeta_i)$ and $(\eta_i^*, \zeta_i^*)$ are arbitrary sets of functions satisfying $s_\sigma[\eta, \zeta]=0=s_\sigma[\eta^*, \zeta^*]$ we must have

$$
(11.10)\qquad \eta_i(x)\zeta_i^*(x) - \zeta_i(x)\eta_i^*(x)\big|_a^b = 0.
$$

Now the general solution of $s_\sigma[\eta, \zeta]=0$, $(\sigma=1, \cdots, 2n)$, is of the form

$$\eta_j(a) = \xi_\tau \overset{1}{c}_{\tau j}, \qquad \zeta_j(a) = \xi_\tau \overset{1}{d}_{\tau j},$$
$$\eta_j(b) = \xi_\tau \overset{2}{c}_{\tau j}, \qquad \zeta_j(b) = -\xi_\tau \overset{2}{d}_{\tau j}, \tag{11.11}$$

where $(\overset{1}{c}_j, \overset{1}{d}_j, \overset{2}{c}_j, \overset{2}{d}_j) = (\overset{1}{c}_{\tau j}, \overset{1}{d}_{\tau j}, \overset{2}{c}_{\tau j}, \overset{2}{d}_{\tau j})$, $(\tau = 1, \cdots, 2n)$, are $2n$ linearly independent solutions of the system

$$\overset{1}{a}_{\sigma j}\overset{1}{c}_j - \overset{1}{b}_{\sigma j}\overset{1}{d}_j + \overset{2}{a}_{\sigma j}\overset{2}{c}_j - \overset{2}{b}_{\sigma j}\overset{2}{d}_j = 0, \qquad \sigma = 1, \cdots, 2n.$$

Corresponding to $\xi \equiv (\xi_\sigma)$, $\xi^* \equiv (\xi^*_\sigma)$, determine the end values of $(\eta_i, \zeta_i)$ and $(\eta_i^*, \zeta_i^*)$ by equations (11.11). Because of the arbitrariness of $\xi$ and $\xi^*$ relation (11.10) then implies

$$\overset{1}{d}_{\sigma j}\overset{1}{c}_{\tau j} - \overset{1}{c}_{\sigma j}\overset{1}{d}_{\tau j} + \overset{2}{d}_{\sigma j}\overset{2}{c}_{\tau j} - \overset{2}{c}_{\sigma j}\overset{2}{d}_{\tau j} = 0, \qquad \sigma, \tau = 1, \cdots, 2n, \tag{11.12}$$

whence it follows that there is a nonsingular matrix $\|E_{\sigma\tau}\|$ such that

$$\overset{1}{d}_{\sigma j} = E_{\sigma\tau}\overset{1}{a}_{\tau j}, \qquad \overset{1}{c}_{\sigma j} = E_{\sigma\tau}\overset{1}{b}_{\tau j}, \qquad \overset{2}{d}_{\sigma j} = E_{\sigma\tau}\overset{2}{a}_{\tau j}, \qquad \overset{2}{c}_{\sigma j} = E_{\sigma\tau}\overset{2}{b}_{\tau j}.$$

Consequently, writing $\xi_\sigma E_{\sigma\tau} = e_\tau$, a set $(\eta_i, \zeta_i)$ is seen to satisfy (11.9) if and only if there are constants $(e_\tau)$ such that

$$\eta_j(a) = e_\tau \overset{1}{b}_{\tau j}, \qquad \zeta_j(a) = e_\tau \overset{1}{a}_{\tau j},$$
$$\eta_j(b) = e_\tau \overset{2}{b}_{\tau j}, \qquad \zeta_j(b) = -e_\tau \overset{2}{a}_{\tau j}. \tag{11.11'}$$

Either from (11.12), or from substitution of (11.11′) in (11.10), it follows that the $2n \times 2n$ matrix

$$\|k_{\sigma\tau}\| \equiv \|\overset{1}{a}_{\sigma j}\overset{1}{b}_{\tau j} + \overset{2}{a}_{\sigma j}\overset{2}{b}_{\tau j}\| \tag{11.13}$$

is symmetric. Now if the $2n \times 2n$ matrix $\|\overset{1}{b}_{\sigma j}, \overset{2}{b}_{\sigma j}\|$ is of rank $2n - p$, denote by $r \equiv (r_\sigma) \equiv (r_{\gamma\sigma})$, $(\gamma = 1, \cdots, p)$, a set of $p$ linearly independent solutions of the equations

$$r_\sigma \overset{1}{b}_{\sigma j} = 0, \qquad r_\sigma \overset{2}{b}_{\sigma j} = 0, \qquad j = 1, \cdots, n.$$

If $(\eta_i, \zeta_i)$ satisfies (11.9), then $\eta_j(a)$, $\eta_j(b)$ must satisfy

$$\Psi_\gamma \equiv \Psi_{\gamma; ja}\eta_j(a) + \Psi_{\gamma; jb}\eta_j(b) = 0, \qquad \gamma = 1, \cdots, p, \tag{11.14}$$

where $\Psi_{\gamma;ja} = r_{\gamma\sigma}\overset{1}{a}_{\sigma j}$, $\Psi_{\gamma;jb} = r_{\gamma\sigma}\overset{2}{a}_{\sigma j}$. Since $s_\sigma[\eta, \zeta]$ are independent linear forms, the $p$ conditions (11.14) are seen to be also linearly independent.

If $p = 2n$, the problem (11.1) is then seen to be equivalent to one of the sort studied in §10 with $2\omega$ defined as the integrand of (11.4), the auxiliary differential equations $\Phi_\alpha = 0$ and boundary conditions determined by (11.5) and (11.14), respectively, the quadratic form $Q \equiv 0$, while the matrix $\mathfrak{K}(x)$ is defined as

$$\mathfrak{K}(x) \equiv \begin{Vmatrix} -k(x) & 0_\beta \\ 0_\alpha & 0_{\alpha\beta} \end{Vmatrix}, \qquad \alpha, \beta = 1, \cdots, n-1. \tag{11.15}$$

In general, it is to be noted that $b^1_{\sigma j}\Psi_{\gamma;ja}+b^2_{\sigma j}\Psi_{\gamma;jb}=0$, $(\sigma=1, \cdots, 2n;$ $\gamma=1, \cdots, p)$, and since the linear forms of (11.14) are independent and $\|b^1_{\sigma j}, b^2_{\sigma j}\|$ is of rank $2n-p$ it follows that if $b^1_{\sigma j}w_{ja}+b^2_{\sigma j}w_{jb}=0$, $(\sigma=1, \cdots, 2n)$, then there must exist constants $d_\gamma$ such that $w_{ja}=d_\gamma\Psi_{\gamma;ja}$, $w_{jb}=d_\gamma\Psi_{\gamma;jb}$, $(j=1, \cdots, n)$. Moreover, for $(e_\tau)$ related to the end values of $(\eta_i, \zeta_i)$ by (11.11′) we have

$$\overset{1}{a}_{\sigma j}\eta_j(a) + \overset{2}{a}_{\sigma j}\eta_j(b) = k_{\sigma\tau}e_\tau. \tag{11.16}$$

Since $r\equiv(r_{\gamma\sigma})$, $(\gamma=1, \cdots, p)$, satisfies $k_{\sigma\tau}r_\tau=0$, the rank of $\|k_{\sigma\tau}\|$ does not exceed $2n-p$. If this matrix is of rank $2n-q$, denote by $\rho\equiv(\rho_\tau)\equiv(\rho_{\nu\tau})$, $(\nu=1, \cdots, q)$, sets orthonormal in the sense that $\rho_{\nu\tau}\rho_{\chi\tau}=\delta_{\nu\chi}$, $(\nu, \chi=1, \cdots, q)$, and satisfying $k_{\sigma\tau}\rho_\tau=0$, $(\sigma=1, \cdots, 2n)$. Then the matrix

$$\begin{Vmatrix} k_{\sigma\tau} & \rho_{\chi\sigma} \\ \rho_{\nu\tau} & 0_{\nu\chi} \end{Vmatrix}$$

is nonsingular, and its reciprocal is a symmetric matrix of the form

$$\begin{Vmatrix} h_{\sigma\tau} & \rho_{\chi\sigma} \\ \rho_{\nu\tau} & 0_{\nu\chi} \end{Vmatrix}.$$

If now $(\eta_i, \zeta_i)$ satisfies (11.9), and $(e_\tau)$ is determined by (11.11′), it follows from (11.16) that there exist constants $t_\nu$, $(\nu=1, \cdots, q)$, such that

$$e_\sigma = h_{\sigma\tau}[\overset{1}{a}_{\tau j}\eta_j(a) + \overset{2}{a}_{\tau j}\eta_j(b)] + t_\nu\rho_{\nu\sigma}. \tag{11.17}$$

Writing

$$2Q[\eta] \equiv [\overset{1}{a}_{\sigma i}\eta_i(a) + \overset{2}{a}_{\sigma i}\eta_i(b)]h_{\sigma\tau}[\overset{1}{a}_{\tau j}\eta_j(a) + \overset{2}{a}_{\tau j}\eta_j(b)], \tag{11.18}$$

it then follows from (11.11′) and (11.17) that

$$\begin{aligned} &Q_{ia}[\eta] + t_\nu\rho_{\nu\tau}\overset{1}{a}_{\tau i} - \zeta_i(a) = 0, \\ &Q_{ib}[\eta] + t_\nu\rho_{\nu\tau}\overset{2}{a}_{\tau i} + \zeta_i(b) = 0, \qquad i = 1, \cdots, n. \end{aligned} \tag{11.19}$$

On the other hand, since

$$\overset{1}{b}_{\sigma j}(t_\nu\rho_{\nu\tau}\overset{1}{a}_{\tau j}) + \overset{2}{b}_{\sigma j}(t_\nu\rho_{\nu\tau}\overset{2}{a}_{\tau j}) = t_\nu(\rho_{\nu\tau}k_{\tau\sigma}) = 0,$$

there exist constants $d_\gamma$ such that $t_\nu\rho_{\nu\tau}a^1_{\tau j}=d_\gamma\Psi_{\gamma;ja}$, $t_\nu\rho_{\nu\tau}a^2_{\tau j}=d_\gamma\Psi_{\gamma;jb}$. It thus follows that (11.1) is equivalent to a boundary value problem of the type treated in the preceding section with $2\omega$ defined as the integrand of (11.4), the auxiliary differential equations and end conditions defined by (11.5) and (11.14), respectively, the quadratic form $Q$ of (11.18), and $\mathfrak{K}(x)$ given in (11.15).

Finally, if $u$ is of class $C^{(2n)}$ on $ab$ and $(\eta_i, \zeta_i)$ are defined by (11.7), it is readily seen that $u\mathfrak{L}[u] \equiv 2\omega(x, \eta, \eta') - (\eta_i \zeta_i)'$ and

$$\int_a^b u\mathfrak{L}[u]\, dx \equiv \int_a^b 2\omega(x, \eta, \eta')\, dx - \eta_i\zeta_i \Big|_a^b = \int_a^b 2\omega(x, \eta, \eta')\, dx + 2Q[\eta]$$

whenever $U_\sigma[u]=0$. In particular, the hypotheses of Kamke on (11.1) are seen to imply that the above defined equivalent problem is $H$-definitely self-adjoint in the sense of §2. Actually, a self-adjoint system (11.1) is $H$-definitely self-adjoint if $\lambda=0$ is not a characteristic value, and the functional (11.2) is non-negative for arbitrary functions $u$ of class $C^{(2n)}$ satisfying with a continuous function $g(x)$ the system $L[u]=k(x)g(x)$, $U_\sigma[u]=0$. In case $k(x)$ vanishes or changes sign on $ab$ this condition is slightly weaker than that used by Kamke.

In conclusion, it is to be remarked that once the symmetry of $\|k_{\sigma\tau}\|$ is established, the existence of linear forms $\Psi_\gamma$ and a quadratic form $Q$ such that the boundary conditions $s_\sigma[\eta, \zeta]=0$ reduce to (11.14), (11.19) has been proved by Hu [4, pp. 380–382]. The above presentation, however, determines more explicitly the form of the $\Psi_\gamma$ and $Q$ in terms of the coefficients of the forms $s_\sigma[\eta, \zeta]$.

12. ***H*-definitely self-conjugate adjoint systems.** In the preceding sections we have been concerned with a system (2.1) involving real-valued coefficients. However, the notion of $H$-definite self-adjointness may be extended to a system (2.1) whose coefficients are complex-valued in a manner previously presented by Reid [11] for extending the notion of definite self-adjointness to such a system.

In the following we shall therefore suppose that the elements of $A(x)$, $B(x)$ are complex-valued continuous functions of the real variable $x$ on $ab$, and that the coefficient matrices $M$ and $N$ of the linearly independent boundary conditions $s_i[y]=0$ have complex-valued elements. If $K\equiv\|K_{ij}\|$, then we shall denote by $\bar{K}$ the matrix $\|\bar{K}_{ij}\|$ whose elements are the complex conjugates of the corresponding elements of $K$; moreover, $K^*$ shall denote the conjugate transpose matrix $\|\bar{K}_{ji}\|$. As in Reid [11] we shall also consider the system

$$(12.1) \qquad u' + u\bar{A} = -\lambda u\bar{B}, \qquad \bar{t}[u] \equiv u(a)\bar{P} + u(b)\bar{Q} = 0,$$

where $P$ and $Q$ are the matrices occurring in the boundary conditions of the adjoint system (2.2). System (12.1) is termed the *conjugate adjoint* of (2.1). The system (2.1) is said to be *self-conjugate adjoint* with the matrix $T$ if it is equivalent to (12.1) under the transformation $u=T(x)y$, where the elements of $T(x)$ are complex-valued functions which are of class $C^1$ on $ab$, and $T$ is non-singular on this interval. It follows (Reid [11, Theorem 2.1]) that (2.1) is self-conjugate adjoint with $T(x)$ if and only if

$$(12.2) \qquad \begin{gathered} TA + A^*T + T' \equiv 0, \qquad TB + B^*T \equiv 0 \quad \text{on} \quad ab, \\ MT^{-1}(a)M^* \equiv NT^{-1}(b)N^*. \end{gathered}$$

We shall now say that (2.1) is *H-definitely self-conjugate adjoint with the matrix T*, or merely *H-definitely self-conjugate adjoint* if:

(i) The system is self-conjugate adjoint with $T$.

(ii) The matrix $S(x) \equiv T^*(x)B(x)$ is hermitian.

(iii) If the linear vector space $L$ be defined as in §2, with the understanding now that the components of $y$ and $g$ are complex-valued, then the functional

$$H[y] \equiv \int_a^b \bar{y}\, T^*\mathcal{L}[y]\, dx,$$

which is readily seen to be real-valued on this space $L$, is positive for arbitrary vectors $y$ of $L$ such that $By \not\equiv 0$ on $ab$.

(iv) There exists no nonidentically vanishing solution $y$ of $\mathcal{L}[y] = 0$, $s[y] = 0$ such that $By \equiv 0$ on $ab$.

THEOREM 12.1. *All the characteristic values of an H-definitely self-conjugate adjoint system* (2.1) *are real.*

For if $y$ were a characteristic solution of an $H$-definitely self-conjugate adjoint system corresponding to a non-real characteristic value $\lambda$, it would follow as in the proof of Theorem 3.1 of Reid [11] that $\int_a^b \bar{y}Sy\, dx = 0$. Since for such a characteristic solution we have $H[y] = \lambda\int_a^b \bar{y}Sy\, dx$, it then ensues that $H[y] = 0$. Because of the above condition (iii) it would then follow that $By \equiv 0$ on $ab$, which is impossible for a characteristic solution by condition (iv). Hence all the characteristic values of such a system are real.

Once this result is obtained, the consideration of the existence of characteristic values and related expansion theorems for an $H$-definitely self-conjugate adjoint system (2.1) is reducible to the same consideration for an associated $H$-definitely self-adjoint system with real coefficients. Since this reduction is attained by the same device of separating real and pure imaginary parts of (2.1) for *real* values of $\lambda$ as used in Reid [11], the details of the reduction will be left to the reader.

In a general discussion of boundary value problems one might very well start with a system of the form (2.1) whose coefficients are complex-valued, which satisfies the above conditions (i), (ii), (iv) and the following alternative to the above condition (iii):

(iii)* If the linear vector space $L$ be defined as in §2, with the understanding that the components of $y$ and $g$ are complex-valued, then there exist real constants $\alpha$ and $\beta$ not both zero and such that the functional

$$(12.3) \qquad \int_a^b \bar{y}\, T^*(\alpha L[y] + \beta By)\, dx,$$

which is readily seen to be real on $L$, is positive for arbitrary vectors $y$ of $L$ such that $By \not\equiv 0$ on $ab$.

If a system (2.1) satisfies (i), (ii), (iii)* and (iv), and has $\alpha \neq 0$ in (12.3), this system may be reduced to an $H$-definitely self-conjugate adjoint system by a linear change of parameter and the possibly needed change of replacing $T$ by $-T$. If for such a system we have $\alpha = 0$, then the system obtained is somewhat more general than a definitely self-conjugate adjoint system; for such a problem, however, one is still able by the usual method of proof to establish the reality of characteristic values, the equality of index and multiplicity of its characteristic values, and a completeness property of the totality of characteristic solutions similar to that proved by Bliss for definitely self-adjoint systems (see Bliss [2, Theorem 2.3 and its Corollaries]). In a recent course on boundary value problems the author has followed this order of presentation. For the purpose of publication of new results, however, the above separate treatment of $H$-definitely self-adjoint systems seems desirable, since by this procedure one is able on various occasions to utilize readily certain results that have previously been established by Bliss and the author.

## Bibliography

1. G. A. Bliss, *A boundary value problem for a system of ordinary linear differential equations of the first order*, these Transactions, vol. 28 (1926), pp. 561–589.

2. ———, *Definitely self-adjoint boundary value problems*, these Transactions, vol. 44 (1938), pp. 413–428.

3. E. Goursat, *Cours d'Analyse Mathématique*, vol. 3, Paris, 1927.

4. K. S. Hu, *The problem of Bolza and its accessory boundary value problem*, Contributions to the Calculus of Variations, 1931–1932, The University of Chicago Press, pp. 361–443.

5. E. L. Ince, *Ordinary Differential Equations*, 1927.

6. E. Kamke, *Über die definiten selbstadjungierten Eigenwertaufgaben bei gewöhnlichen linearen Differentialgleichungen*. I, Mathematische Zeitschrift, vol. 45 (1939), pp. 759–787; II, ibid., vol. 46 (1940), pp. 231–250; III, ibid., vol. 46 (1940), pp. 251–286.

7. M. Krein, *Sur les opérateurs différentiels autoadjoints et leurs fonctions de Green symétriques*, Recueil Mathématique, vol. 2 (44) (1937), pp. 1023–1070.

8. J. Mercer, *Functions of positive and negative type, and the connection with the theory of integral equations*, Philosophical Transactions of the Royal Society, vol. 209 A (1909), pp. 415–446.

9. W. T. Reid, *A boundary value problem associated with the calculus of variations*, American Journal of Mathematics, vol. 54 (1932), pp. 769–790.

10. ———, *A system of ordinary linear differential equations with two-point boundary conditions*, these Transactions, vol. 44 (1938), pp. 508–521.

11. ———, *Some remarks on linear differential systems*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 414–419.

University of Chicago,
Chicago, Ill.

# BOUNDED UNIVALENT FUNCTIONS

BY

RAPHAEL M. ROBINSON

1. **Introduction.** We shall consider the class of functions $f(z)$ which are regular and univalent for $|z|<1$, with $|f(z)|<1$ there, and with $f(0)=0$. For any fixed $z_0 \neq 0$ in the unit circle, we use the abbreviations

$$(1) \qquad a = |f'(0)|, \quad b = |z_0|, \quad c = |f(z_0)|, \quad d = |f'(z_0)|.$$

We are concerned in this paper with the inequalities relating $a, b, c, d$. It may be noted that these relations are not affected if we impose the condition $f'(0)>0$, which we shall do.

The four quantities (1) are restricted individually only by

$$(2) \qquad 0 < a \leqq 1, \quad 0 < b < 1, \quad 0 < c < 1, \quad 0 < d.$$

If $a=1$, then $f(z)=z$, hence $d=1$ and $c=b$. This trivial case will be excluded below where convenient. If $a$ has any other given value, then it is easily seen that no restriction is placed on any one of the other quantities. Between $b$ and $c$ the only relation is $c \leqq b$; the equality $c=b$ holds only if $a=1$. *The relations among the quantities of each of the sets* $(b, d)$, $(c, d)$, $(a, b, c)$, $(b, c, d)$, *and* $(a, b, c, d)$ *are considered in* §5. *The relations between* $(a, b, d)$ *are considered in* §6, *and those between* $(a, c, d)$ *in* §7. Thus all subsets of the four quantities are considered. It should be pointed out that the determination of the inequalities satisfied by the four quantities by no means completes the solution, since one of the main difficulties is that of eliminating one of the quantities in order to find the inequalities satisfied by three of them.

*All of the inequalities which we obtain will be sharp*; that is, in each case there is an extremal function for which the inequality becomes an equality. But in general we shall not go beyond the mere existence of such an extremal function.

Finally, there is an appendix (§8) on unbounded univalent functions. *Let* $F(z)$ *be regular and univalent for* $|z|<1$, *and suppose* $F(0)=0$, $F'(0)=1$. *The relations between* $|z_0|$, $|F(z_0)|$, *and* $|F'(z_0)|$ *are discussed in detail.* One result, which may seem surprising, will be mentioned here: *If* $|F(z_0)| \leqq 1/4$ *then*

$$|F'(z_0)| \leqq 1 + 3\cdot 2^{-3/2} = 2.06\cdots,$$

*but if* $|F(z_0)|$ *has a prescribed value greater than* $1/4$, *no upper bound for* $|F'(z_0)|$ *can be given.* The results in the appendix are obtained as limiting cases of results for bounded univalent functions, but without using §6 and §7.

---

Presented to the Society, November 22, 1941; received by the editors November 13, 1941.

2. **The method of Löwner.** If, from our class of bounded univalent functions, a subclass is chosen by means of which any function of the given class can be uniformly approximated in the interior of the unit circle, then the inequalities between any of the quantities $a, b, c, d$ are the same for the subclass as for the whole class, except perhaps as regards the possibility of equality signs holding.

According to Löwner([1]), we may choose the subclass as the class of functions $f(z)$ to which a function $f(z, t)$ can be found, with the following properties. There is a number $I>0$ such that $f(z, t)$ is continuous for $|z|<1$ and $0\leqq t\leqq I$, and is a regular function of $z$ for each fixed $t$. The boundary conditions

$$f(z, 0) = z, \qquad f(z, I) = f(z) \tag{3}$$

are satisfied, so that $f(z)$ may be regarded as having been obtained from the identity by continuous variation. The rate at which this variation takes place is governed by

$$f'(0, t) = e^{-t}, \tag{4}$$

where the prime denotes differentiation with respect to $z$. Finally, there is a continuous function $\kappa(t)$ with $|\kappa(t)|=1$, such that $f(z, t)$ satisfies the differential equation

$$\frac{\partial f(z, t)}{\partial t} = -f(z, t)\frac{1+\kappa(t)f(z, t)}{1-\kappa(t)f(z, t)}. \tag{5}$$

It is also permissible, and more convenient for us, to allow $\kappa(t)$ to be a piece-wise continuous function; we then understand that (5) is to hold except at the points of discontinuity of $\kappa(t)$, and similarly below. This weaker condition on $\kappa(t)$ means that we are choosing a larger subclass from the class of bounded univalent functions. The advantage of this is that extremal functions for all of our inequalities are then brought within the subclass.

Another fact which is important for us is that (5) can be solved for any given $\kappa(t)$ satisfying the conditions mentioned, so that there is a one-to-one correspondence between such functions $\kappa(t)$ and functions $f(z)$ of our subclass.

From (3) and (4) we see that

$$a = e^{-I}. \tag{6}$$

3. **The integrals $I$ and $J$.** From (5) we readily obtain

$$\frac{d}{dt}\log|f(z_0, t)| = -\frac{1-|f(z_0, t)|^2}{|1-\kappa(t)f(z_0, t)|^2}. \tag{7}$$

---

([1]) K. Löwner, *Untersuchungen über schlichte konforme Abbildungen des Einheitskreises*, Mathematische Annalen, vol. 89 (1923), pp. 103–121.

If we put

$$s = |f(z_0, t)|, \tag{8}$$

then $s$ decreases from $b$ to $c$ while $t$ increases from 0 to $I$. Any function of $t$ in the interval $0 \leqq t \leqq I$ may be regarded as a function of $s$ in the interval $c \leqq s \leqq b$. In particular, we put

$$\kappa(t)f(z_0, t) = \eta(s)s, \tag{9}$$

so that $\eta(s)$ is a piece-wise continuous function with $|\eta(s)| = 1$. Then (7) takes the form

$$\frac{ds}{dt} = -\frac{s}{H(s)}, \tag{10}$$

where

$$H(s) = \frac{|1 - \eta(s)s|^2}{1 - s^2}. \tag{11}$$

We have evidently

$$\frac{1 - s}{1 + s} \leqq H(s) \leqq \frac{1 + s}{1 - s}. \tag{12}$$

We show now that $\eta(s)$ is an arbitrary piece-wise continuous function with $|\eta(s)| = 1$. If any such $\eta(s)$ is given in the interval $c \leqq s \leqq b$, then we first determine $H(s)$ from (11), and then from (10) and the fact that $t = 0$ for $s = b$, we find that

$$t = \int_s^b H(\xi) d\xi/\xi, \tag{13}$$

and in particular that

$$I = \int_c^b H(s) ds/s. \tag{14}$$

Now (13) determines $s$ as a function of $t$ in the interval $0 \leqq t \leqq I$, so that (8) and (9) determine $|f(z_0, t)|$ and $\kappa(t)f(z_0, t)$. From (5) it follows that

$$\frac{d}{dt} \operatorname{amp} f(z_0, t) = -\frac{2\Im[\kappa(t)f(z_0, t)]}{|1 - \kappa(t)f(z_0, t)|^2}. \tag{15}$$

Using the value just found for $\kappa(t)f(z_0, t)$, we can find $\operatorname{amp} f(z_0, t)$, and hence $f(z_0, t)$ itself, the initial condition $f(z_0, 0) = z_0$ being imposed. Finally, knowing $\kappa(t)f(z_0, t)$ and $f(z_0, t)$, we can find $\kappa(t)$ by division; it will be piece-wise continuous and satisfy $|\kappa(t)| = 1$. Now the $\kappa(t)$ and $f(z_0, t)$ determined satisfy (7)

and (15), which are equivalent to (5) with $z=z_0$. From this we see that the $\kappa(t)$ which we have found will in fact lead back to the desired $\eta(s)$.

From the fact that $\eta(s)$ is arbitrary, we see that $H(s)$ is an arbitrary piecewise continuous function satisfying (12). For if any such $H(s)$ is given, we can find an $\eta(s)$ satisfying (11).

It is readily seen that if $f(z, t)$ is a continuous function which is regular in $z$ for each fixed $t$, and such that $f_t(z, t)$ is continuous, then $f_{zt}(z, t)$ and $f_{tz}(z, t)$ exist and are equal. Using this fact, from (5) we find that

$$\frac{d}{dt}\log |f'(z_0, t)| = 1 - 2\frac{\Re\{[1-\kappa(t)f(z_0, t)]^2\}}{|1-\kappa(t)f(z_0, t)|^4}. \tag{16}$$

Expressing the right side in terms of $s$, and using (10), we have

$$\frac{d}{dt}\log |f'(z_0, t)| = \frac{-1}{s(1-s^2)}\left[|1-\eta(s)s|^2 - 2\frac{\Re\{[1-\eta(s)s]^2\}}{|1-\eta(s)s|^2}\right]\frac{ds}{dt}.$$

Integrating from $t=0$ to $t=I$ gives

$$\log d = \int_c^b \frac{|1-\eta(s)s|^2 - 2\cos\{2 \operatorname{amp}[1-\eta(s)s]\}}{s(1-s^2)}\,ds. \tag{17}$$

Now

$$\cos \operatorname{amp}[1-\eta(s)s] = \frac{1-s^2+|1-\eta(s)s|^2}{2|1-\eta(s)s|},$$

so that the numerator of the integrand becomes $2s^2-(1-s^2)/H(s)$, and (17) takes the form

$$d = \frac{1-c^2}{1-b^2}e^{-J}, \tag{18}$$

where

$$J = \int_c^b (1/H(s))ds/s. \tag{19}$$

The problem of finding what values are possible for $a$ and $d$ when $b$ and $c$ are given is thus reduced to finding the relations between the integrals $I$ and $J$. The corresponding values of $a$ and $d$ are then found from (6) and (18).

4. **Relations between $I$ and $J$.** Suppose $0<c<b<1$, and let $I$ and $J$ be defined by (14) and (19), where $H(s)$ is any function satisfying (12) and piecewise continuous in the interval $c\leqq s\leqq b$. In this section, we shall determine the inequalities relating $I$ and $J$. It is clear that the relation between $I$ and $J$ is a symmetric one.

We start by introducing two functions which we shall need in this discussion. For $c\leqq r\leqq b$, let

$$(20)\qquad \begin{aligned} p(r;b,c) &= \int_c^r \frac{1-s}{1+s}\frac{ds}{s} + \int_r^b \frac{1-r}{1+r}\frac{ds}{s},\\ q(r;b,c) &= \int_c^r \frac{1+s}{1-s}\frac{ds}{s} + \int_r^b \frac{1+r}{1-r}\frac{ds}{s}. \end{aligned}$$

Evaluating the integrals gives

$$(21)\qquad \begin{aligned} p(r;b,c) &= \log\left[\frac{r}{(1+r)^2}:\frac{c}{(1+c)^2}\right] + \frac{1-r}{1+r}\log\frac{b}{r},\\ q(r;b,c) &= \log\left[\frac{r}{(1-r)^2}:\frac{c}{(1-c)^2}\right] + \frac{1+r}{1-r}\log\frac{b}{r}; \end{aligned}$$

in particular we have

$$(22)\qquad \begin{aligned} p(b;b,c) &= \log\left[\frac{b}{(1+b)^2}:\frac{c}{(1+c)^2}\right], \quad p(c;b,c) = \frac{1-c}{1+c}\log\frac{b}{c},\\ q(b;b,c) &= \log\left[\frac{b}{(1-b)^2}:\frac{c}{(1-c)^2}\right], \quad q(c;b,c) = \frac{1+c}{1-c}\log\frac{b}{c}. \end{aligned}$$

It is clear from (20) that $p(r;b,c)$ is a decreasing function of $r$, and $q(r;b,c)$ an increasing function. Since $p(r;b,c)<q(r;b,c)$, we have in particular

$$(23)\qquad p(b;b,c) < p(c;b,c) < q(c;b,c) < q(b;b,c).$$

It is clear first of all that individually $I$ and $J$ are restricted only by the conditions

$$(24)\qquad p(b;b,c) \leqq I \leqq q(b;b,c),$$

$$(25)\qquad p(b;b,c) \leqq J \leqq q(b;b,c).$$

To find the largest possible value of $J$ when $I$ is given, we note that $H(s)+1/H(s) \leqq (1-s)/(1+s)+(1+s)/(1-s)$, and hence

$$(26)\qquad I + J \leqq p(b;b,c) + q(b;b,c).$$

The equality is attained for any piece-wise continuous function $H(s)$ which is equal to $(1-s)/(1+s)$ in some subintervals, and to $(1+s)/(1-s)$ in others. Since these are the values of $H(s)$ which give $I$ its smallest and largest values, we see that any possible value of $I$ can be obtained in this way. Hence for any given $I$, the largest possible $J$ is determined from (26).

It remains to find the smallest possible $J$ for a given $I$. Let $k$ be any positive constant, and consider

$$(27)\qquad I + k^2J = \int_c^b \left[H(s) + \frac{k^2}{H(s)}\right]\frac{ds}{s}.$$

The integrand is smallest when $H(s)=k$; but this may not be compatible with (12). It is clear that the integral will be minimized if we keep $H(s)$ as near to $k$ as possible. If $k$ is sufficiently small, then this $H(s)$ is always equal to $(1-s)/(1+s)$; and if $k$ is sufficiently large, then $H(s)=(1+s)/(1-s)$. Hence for a suitable value of $k$, $I$ may be given any possible value. Thus the minimum $J$ for any given $I$ is obtained for $H(s)$ as near to some constant $k$ as possible.

To formulate the lower bound for $J$, we distinguish three cases:

$$\begin{array}{ll} \text{Case } p. & p(b;\, b,\, c) \leqq I \leqq p(c;\, b,\, c). \\ \text{Case } o. & p(c;\, b,\, c) \leqq I \leqq q(c;\, b,\, c). \\ \text{Case } q. & q(c;\, b,\, c) \leqq I \leqq q(b;\, b,\, c). \end{array} \tag{28}$$

We note first that $p(r;\, b,\, c)$ is given by (14) where $H(s)$ is as near to $(1-r)/(1+r)$ as possible. Hence in Case $p$, we determine $r$ so that $p(r;\, b,\, c)=I$, and then $J\geqq q(r;\, b,\, c)$. Similarly, in Case $q$, $J\geqq p(r;\, b,\, c)$, where $q(r;\, b,\, c)=I$. In Case $o$, $H(s)$ may be equal to a constant $k$ throughout the interval, and hence $J$ is minimized in this way. Hence for $I=k \log b/c$, the minimum $J$ is $(1/k) \log b/c$, so that $J\geqq(\log b/c)^2/I$. The three cases may be combined in the form

$$J \geqq L(I;\, b,\, c) \tag{29}$$

where

$$L(I;\, b,\, c) = \begin{cases} q(r;\, b,\, c), \text{ where } p(r;\, b,\, c) = I, & \text{in Case } p, \\ (\log b/c)^2/I, & \text{in Case } o, \\ p(r;\, b,\, c), \text{ where } q(r;\, b,\, c) = I, & \text{in Case } q. \end{cases} \tag{30}$$

5. **The simpler cases of the problem.** In this section we obtain the inequalities among each set of quantities chosen from $(a, b, c, d)$, except for the trivial cases treated in the introduction, and the cases $(a, b, d)$ and $(a, c, d)$, which have separate sections devoted to them. However, partial results for those two cases are given here.

**Relations between** $a$, $b$, $c$. These were first obtained by Pick[2]; more recently, Golusin[3] derived them, using the method of Löwner. From (24) we have, for $c<b$,

$$\frac{c}{(1-c)^2} : \frac{b}{(1-b)^2} \leqq a \leqq \frac{c}{(1+c)^2} : \frac{b}{(1+b)^2}. \tag{31}$$

---

(2) G. Pick, *Über die konforme Abbildung eines Kreises auf ein schlichtes und zugleich beschränktes Gebiet*, Sitzungsberichte Akademie der Wissenschaften, Vienna, vol. 126 (1917), pp. 247–263.

(3) G. M. Golusin, *Über die Verzerrungssätze der schlichten konformen Abbildungen* (Russian with German summary), Matematicheskii Sbornik (Recueil Mathématique), vol. 43 (1936), pp. 127–135.

For $c=b$, (31) gives $a=1$, which is correct. Thus for any given $b$ and $c$, with $c\leqq b$, the value of $a$ is restricted only by (31). It may be noted that (31) itself implies that $c\leqq b$, so that it is in fact the only relation between $a, b, c$.

The bounds in (31) are attained for the function $w=f(z)$ defined for $|z|<1$ by

$$\frac{w}{(1-w)^2}=\frac{az}{(1-z)^2}, \qquad |w|<1. \tag{32}$$

This function maps $|z|<1$ on $|w|<1$ with a slit along the negative real axis. The lower and upper bounds are attained for $z_0>0$ and $z_0<0$, respectively.

**Relations between** $b, c, d$. From (25) we have

$$\frac{c}{(1-c)^2}:\frac{b}{(1-b)^2}\leqq d\,\frac{1-b^2}{1-c^2}\leqq\frac{c}{(1+c)^2}:\frac{b}{(1+b)^2}. \tag{33}$$

This inequality could also be obtained from (31) by making linear transformations of $|z|<1$ and $|w|<1$ into themselves, in such a way that 0 and $z_0$ are interchanged in the $z$-plane, and 0 and $f(z_0)$ in the $w$-plane. From (33) we find that

$$\frac{c(1+c)}{1-c}:\frac{b(1+b)}{1-b}\leqq d\leqq\frac{c(1-c)}{1+c}:\frac{b(1-b)}{1+b}. \tag{34}$$

This is the only relation between $b, c, d$. The bounds are attained for the function (32), for $z_0<0$ and $z_0>0$, respectively.

**Relations between** $b$ **and** $d$. This case may be solved by seeing what bounds are given for $d$ by (34) when $b$ is given but $c$ is not. The only restriction on $c$ is that $0<c\leqq b$. Letting $c\rightarrow 0$, we see that there is no positive lower bound for $d$. On the other hand, the right side of (34) is bounded, and in fact has its largest value for $c=2^{1/2}-1$; if this is not within the allowed interval, then the largest possible value is at $c=b$. From this we find that

$$d\leqq\begin{cases}1 & \text{if } 0<b\leqq 2^{1/2}-1,\\ (3-2^{3/2})\dfrac{1+b}{b(1-b)} & \text{if } 2^{1/2}-1\leqq b<1.\end{cases} \tag{35}$$

This is the only relation between $b$ and $d$. The equality sign holds for the identity in the first case, and for the function (32) with $a$ chosen so that $c=2^{1/2}-1$ in the second. Dieudonné[4] has shown that the first part of (35) holds for bounded functions which are not supposed univalent.

**Relations between** $c$ **and** $d$. Letting $b\rightarrow 1$ in (34), we see that there is no restriction on the value of $d$ if only $c$ is given.

---

[4] J. Dieudonné, *Recherches sur quelques problèmes relatifs aux polynômes et aux fonctions bornées d'une variable complexe*, Annales de l'École Normale, vol. (3) 48 (1931), p. 352.

**Relations between $a, b, d$ (partial results).** From (31) and (34) it is possible to find the smallest value of $d$ for given $a$ and $b$, and in some cases also the largest $d$.

If $a$ and $b$ are given, we may determine the smallest possible value of $c$ from the right side of (31), and then the smallest possible $d$ for this $c$ and the given $b$ from the left side of (34). Since the two functions of $c$ involved in the bounds are increasing, and since the equality signs in the two cases are attained together, we obtain in this way the best lower bound for $d$ in terms of $a$ and $b$:

$$d \geqq \frac{c(1+c)}{1-c} : \frac{b(1+b)}{1-b}, \tag{36}$$

where $c$ is determined from

$$\frac{c}{(1+c)^2} = \frac{ab}{(1+b)^2}. \tag{37}$$

The equality in (36) is attained for (32) with $z_0 < 0$.

Similarly, the equality signs on the left side of (31) and the right side of (34) are attained together. But the function $c(1-c)/(1+c)$, which occurs on the right side of (34), is increasing only for $c \leqq 2^{1/2}-1$, so that we can draw the conclusion

$$d \leqq \frac{c(1-c)}{1+c} : \frac{b(1-b)}{1+b}, \tag{38}$$

where $c$ is determined from

$$\frac{c}{(1-c)^2} = \frac{ab}{(1-b)^2}, \tag{39}$$

only if that value of $c$ is not greater than $2^{1/2}-1$. In this case, the equality will be attained for (32) with $z_0 > 0$. It is clear that we obtain in this way the best upper bound for $d$ in terms of $a$ and $b$ if $b \leqq 2^{1/2}-1$. We shall show in §6 that the same is true whenever $b \leqq 1/2$ and in some other cases, but not in all cases.

**Relations between $a, c, d$ (partial results).** We try to find bounds for $d$ in terms of $a$ and $c$ from (31) and (34). If

$$a > \frac{4c}{(1+c)^2}, \tag{40}$$

then the right side of (31) determines a largest value possible for $b$, and then with this $b$ and the given $c$, a lower bound for $d$ is determined from the left side of (34). This bound is the best possible. It is given by (36) with $b$ determined from (37), and is attained for (32) with $z_0 < 0$. On the other hand, if

(40) is not satisfied, then $b$ is permitted values arbitrarily close to 1. It is easily shown that no positive lower bound exists for $d$ in this case. For example, it is sufficient to consider the functions $w=f(z)$ which map $|z|<1$ on $|w|<1$ with a slit from $-1$ nearly to $-c$ and a slit on the positive real axis long enough to give $a$ the proper value, $z_0$ being chosen so that $f(z_0)=-c$.

We can draw the conclusion that an upper bound for $d$ is given by (38) with $b$ determined from (39) only if $b(1-b)/(1+b)$ has its smallest value when $b$ has its smallest possible value; if this is true, then the maximum value of $d$ is attained for (32) with $z_0>0$. The condition is certainly satisfied if for the given values of $a$ and $c$ one has necessarily $b\leqq 2^{1/2}-1$; this is true at least for $a$ near 1 and $c$ near 0. More generally, if we denote the smallest possible value of $b$ by $b_{\min}$, and suppose that $b$ has a largest possible value $b_{\max}$, then the condition is satisfied if $b(1-b)/(1+b)$ is not larger at $b_{\min}$ than at $b_{\max}$. This condition reduces to

$$(1 + b_{\min})(1 + b_{\max}) \leqq 2,$$

which is the best result obtainable by the present method. We shall show in §7 that the conclusion holds if and only if

$$b_{\min} \leqq 1/2, \qquad b_{\max} \leqq 1, \tag{41}$$

where the second inequality is to be interpreted to mean: either $a$ and $c$ have such values that $b_{\max}<1$, or are limits of such values.

**Remark on the hyperbolic expansion factor.** We may interpret the expression $d(1-b^2)/(1-c^2)$, which occurs in (33), as the expansion factor for the mapping $w=f(z)$, when the metric of hyperbolic geometry is introduced in $|z|<1$ and $|w|<1$. By a similar argument to that used above, we find that *no matter which two of the three quantities $a$, $b$, $c$ are given, the hyperbolic expansion is minimized for* (32) *with $z_0<0$, and maximized for* (32) *with $z_0>0$.* Only in case $a$ and $c$ are given, not satisfying (40), and we are seeking to minimize the hyperbolic expansion, is it impossible to satisfy the necessary conditions, $z_0<0$ and $|f(z_0)|=c$. But in this case we know that $d$ has no positive lower bound, and a fortiori the same is true of the hyperbolic expansion. It may also be noted that the conclusion that the hyperbolic expansion has its extreme values for (32) is weaker than the same conclusion about $d$, and hence follows from this when this is true. The bounds for the hyperbolic expansion when $a$ and $b$ are given were found by Pick(8).

So far in this section, we have used from §4 only the trivial results (24) and (25); the rest of the section is used first in considering the relation between all four quantities. It may also be noted that the results of this section so far have depended only on (31) and (33). Since (33) can be deduced from (31), these results can be obtained without using Löwner's method, if we as-

(8) Pick, loc. cit.

sume (31) from the work of Pick. The rest of this section, and the next two sections, depend essentially on Löwner's method.

**Relations between** $a, b, c, d$. On account of (18), we see that the lower bound for $d$ in terms of $a, b, c$ is found from the upper bound for $J$, given in (26). This leads to

$$d \geqq \frac{1}{a} \frac{c^2}{1-c^2} : \frac{b^2}{1-b^2} \cdot \tag{42}$$

It may be verified that this bound is attained for any function $w=f(z)$ mapping $|z|<1$ on $|w|<1$ with slits along the positive and negative real axes; the equality sign in (42) then holds for any positive or negative $z_0$. The lengths of the slits and the value of $z_0$ may be so chosen as to give any desired values to $a, b, c$.

Similarly, the upper bound for $d$ is found from the lower bound for $J$, given by (29). The result may be written in the form

$$\log d \leqq M(I; b, c), \tag{43}$$

where

$$M(I; b, c) = \log \frac{1-c^2}{1-b^2} - L(I; b, c). \tag{44}$$

6. **Relations between** $a, b, d$. We shall obtain the lower and upper bounds for $d$ in terms of $a$ and $b$ by eliminating $c$ from (42) and from (43); this will complete the partial solution given in §5.

The lower bound for $d$ in terms of $a$ and $b$ may be obtained from (42) by substituting the smallest possible value of $c$, which is obtained from (37). This is seen to agree with our previous result, which was (36) with the same value of $c$ substituted.

We turn now to the problem of finding the upper bound for $d$ in terms of $a$ and $b$. We have to maximize $M(I; b, c)$ for all possible values of $c$. Now (44) defines $M(I; b, c)$ in terms of $L(I; b, c)$, which in turn is defined by (30). In (30), different formulas hold in each of the three cases (28). The cases are distinguished according to the interval in which $I$ lies when $b$ and $c$ are given. But now we wish to consider $I$ and $b$ as given, and see in what intervals $c$ must lie in order that each of the cases may hold.

We note first that all four functions $p(b; b, c)$, $p(c; b, c)$, $q(c; b, c)$, $q(b; b, c)$ are decreasing functions of $c$. This is evident from (22) for all the functions but $q(c; b, c)$. For fixed $b$, let $\phi(c) = q(c; b, c)$; then

$$\phi(c) = \frac{1+c}{1-c} \log \frac{b}{c}, \qquad \phi'(c) = \frac{2}{(1-c)^2} \log \frac{b}{c} - \frac{1+c}{c(1-c)} \cdot$$

If we put $\psi(c) = \log c + (1-c^2)/2c$, then the condition $\phi'(c)<0$ reduces to the

form $\psi(c) > \log b$. Now $\psi'(c) = -(1-c)^2/2c^2 < 0$, so that $\psi(c)$ is decreasing. Hence $\psi(c) > \psi(b) > \log b$, as was to be shown.

Thus each of the four functions is strictly monotone, and from (22) it is seen that each decreases from $+\infty$ to 0 as $c$ increases from 0 to $b$. Hence if $a$ and $b$ are given, with $a < 1$, there are unique numbers $c_p$, $\bar{c}_p$, $\bar{c}_q$, $c_q$, between 0 and $b$, which satisfy

$$p(b;\, b,\, c_p) = p(\bar{c}_p;\, b,\, \bar{c}_p) = q(\bar{c}_q;\, b,\, \bar{c}_q) = q(b;\, b,\, c_q) = I. \tag{45}$$

From (23) it is seen that these numbers satisfy the inequalities

$$c_p < \bar{c}_p < \bar{c}_q < c_q; \tag{46}$$

and (24) takes the form $c_p \leqq c \leqq c_q$. The three cases (28) are equivalent to the following:

$$\begin{aligned} &\text{Case } p. \quad c_p \leqq c \leqq \bar{c}_p.\\ &\text{Case } o. \quad \bar{c}_p \leqq c \leqq \bar{c}_q.\\ &\text{Case } q. \quad \bar{c}_q \leqq c \leqq c_q.\end{aligned} \tag{47}$$

We now calculate $L_c(I;\, b,\, c)$ in each of the three cases. In the first place, we have

$$\begin{aligned} p_r(r;\, b,\, c) &= \frac{-2}{(1+r)^2}\log\frac{b}{r}, \qquad & p_c(r;\, b,\, c) &= -\frac{1-c}{c(1+c)},\\ q_r(r;\, b,\, c) &= \frac{2}{(1-r)^2}\log\frac{b}{r}, \qquad & q_c(r;\, b,\, c) &= -\frac{1+c}{c(1-c)}\,. \end{aligned} \tag{48}$$

From these and the definition (30), we have

$$L_c(I;\, b,\, c) = \begin{cases} -\dfrac{1+c}{c(1-c)} - \left(\dfrac{1+r}{1-r}\right)^2 \dfrac{1-c}{c(1+c)} \text{ where } p(r;\, b,\, c) = I & \text{(Case } p),\\[2ex] -\dfrac{2}{cI}\log\dfrac{b}{c} & \text{(Case } o),\\[2ex] -\dfrac{1-c}{c(1+c)} - \left(\dfrac{1-r}{1+r}\right)^2 \dfrac{1+c}{c(1-c)} \text{ where } q(r;\, b,\, c) = I & \text{(Case } q). \end{cases} \tag{49}$$

In particular, the values of $L_c(I;\, b,\, c_p)$ and $L_c(I;\, b,\, c_q)$ are obtained from the first and last parts of (49) by putting $r = b$. Furthermore, we find from (49) that

$$L_c(I;\, b,\, \bar{c}_p) = -\frac{2(1+\bar{c}_p)}{\bar{c}_p(1-\bar{c}_p)}, \qquad L_c(I;\, b,\, \bar{c}_q) = -\frac{2(1-\bar{c}_q)}{\bar{c}_q(1+\bar{c}_q)}; \tag{50}$$

that is, both the left- and right-hand derivatives at $\bar{c}_p$ and $\bar{c}_q$ have these values. Thus $L_c(I;\, b,\, c)$ exists everywhere and is continuous.

It is evident from (49) that $L_c(I; b, c) < 0$ in all cases. Hence from (29) we see that the smallest $J$ for given $a$ and $b$ is obtained for $c = c_q$; we thus verify again that the hyperbolic expansion is maximized in this case. This does not however tell when $d$ itself is largest. For this purpose, we have to maximize $M(I; b, c)$. From the definition (44) we find that

$$(51) \qquad M_c(I; b, c) = -\frac{2c}{1-c^2} - L_c(I; b, c).$$

We must investigate the sign of $M_c(I; b, c)$ in each of the cases (47).

Case $p$. We see at once that $M_c(I; b, c) > 0$, so that the maximum value of $M(I; b, c)$ does not occur in this interval.

Case $o$. We find from (50) that

$$(52) \qquad M_c(I; b, \bar{c}_p) > 0, \quad M_c(I; b, \bar{c}_q) >, =, < 0 \text{ according as } \bar{c}_q <, =, > 1/2.$$

Now $M_c(I; b, c)$ is seen to be decreasing, so that it is negative, if at all, in a subinterval abutting $\bar{c}_q$. Hence $M(I; b, c)$ is monotone increasing in the interval if $\bar{c}_q \leqq 1/2$, while if $\bar{c}_q > 1/2$ it increases to a maximum and then decreases. In the latter case, the maximum is at a point $c > 1/2$. For the condition $\bar{c}_q > 1/2$ is equivalent to $q(1/2; b, 1/2) > I$ or $3 \log 2b > I$; the condition $M_c(I; b, 1/2) > 0$ reduces to the same form if $c = 1/2$ comes in Case $o$, and is trivial if it comes in Case $p$.

Case $q$. The condition $M_c(I; b, c) > 0$ is seen to reduce to

$$(53) \qquad (1 + c)^2 < (1 + r)^2/2r \qquad \text{where } q(r; b, \bar{c}) = I.$$

Since $r$ increases with $c$, we see that (53) is more likely to be true the smaller $c$ is. Hence $M(I; b, c)$ first increases and then decreases, or else is monotone increasing or decreasing. It starts to increase if $\bar{c}_q < 1/2$, and increases throughout the interval if

$$(54) \qquad (1 + c_q)^2 \leqq (1 + b)^2/2b.$$

We note also that (53) is certainly true if $c \leqq 2^{1/2} - 1$, since the right side is more than 2; and it is certainly false if $c \geqq 1/2$, since then $r \geqq c \geqq 1/2$, so that $(1+r)^2/2r \leqq 9/4 \leqq (1+c)^2$.

Putting together the results from the three cases, we see that either $M(I; b, c)$ is monotone increasing in the whole interval $c_p \leqq c \leqq c_q$, or else it first increases and then decreases. Its largest value is at a point $c$ satisfying the following conditions: $c > \bar{c}_p$; either $\bar{c}_q < c < 1/2$, $\bar{c}_q = c = 1/2$, or $\bar{c}_q > c > 1/2$; $c = c_q$ if and only if (54) is true; and if $c \neq c_q$ then $c > 2^{1/2} - 1$.

The conditions involved here may be expressed in terms of $a$ and $b$. In the first place,

$$(55) \qquad \bar{c}_q <, =, > 1/2 \text{ according as } a <, =, > 1/8b^3.$$

Also, (54) with $c_q$ determined as the root of (39) is equivalent to

$$a \leqq \frac{c}{(1-c)^2} : \frac{b}{(1-b)^2} \quad \text{where } (1+c)^2 = \frac{(1+b)^2}{2b}. \tag{56}$$

Substituting the value of $c$, this becomes

$$a \leqq \frac{(1-b)^2[2b(3-2b+3b^2)+(1+b)^3(2b)^{1/2}]}{b(1-6b+b^2)^2}. \tag{57}$$

Using the form (56), we see that this condition is true whenever $b \leqq 1/2$, since then $c \geqq 1/2$; but that is is not true for $b > 1/2$ and $a$ near 1.

We now state the best upper bound for $d$ in the various cases. The cases are distinguished according to the point $c$ where $M(I; b, c)$ is largest.

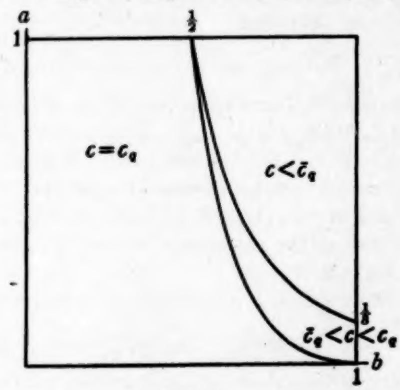


FIG. 1

Case 1. $c \leqq \bar{c}_q$. This is true if $\bar{c}_q \geqq 1/2$, or if $a \geqq 1/8b^3$. The value of $c$ is found from $M_c(I; b, c) = 0$, using the formula of Case $o$. It is easily seen that the bound may be written in the form

$$\log d \leqq \log \frac{1-c^2}{1-b^2} - \frac{c^2}{1-c^2} \log \frac{b}{c} \qquad \text{where } \log \frac{b}{c} = \frac{c^2}{1-c^2} \log \frac{1}{a}. \tag{58}$$

Case 2. $\bar{c}_q < c < c_q$. This is true if $a < 1/8b^3$ and (57) is false. Here

$$\begin{aligned} &\log d \leqq \log \frac{1-c^2}{1-b^2} - p(r; b, c) \\ &\qquad\qquad \text{where } q(r; b, c) = \log \frac{1}{a} \text{ and } (1+c)^2 = \frac{(1+r)^2}{2r}. \end{aligned} \tag{59}$$

Case 3. $c = c_q$. This holds if (57) is true, and in particular if $b \leqq 1/2$. We then have the result (38), which was previously obtained under more restrictive conditions.

The values of $a$ and $b$ for which the various cases hold are shown in Figure 1.

7. **Relations between** $a, c, d$. The lower bound for $d$ in terms of $a$ and $c$ is obtained from (42) by taking $b$ as large as possible, consistent with (31). If (40) is satisfied, then there is a largest value possible for $b$; the required lower bound for $d$ is (42) with $b$ determined from (37), which is seen to be the same as (36) with the same value of $b$ substituted. If (40) is false, then $b$ may be arbitrarily near to 1, and (42) shows that there is no positive lower bound for $d$. Thus our previous results for this case are checked.

We now turn to the problem of finding the upper bound for $d$ in terms of $a$ and $c$. To do this, we have to eliminate $b$ from (43). This turns out to be the most difficult problem of all. In order to prevent this section from being unreasonably long, we shall omit a number of calculations; but some of these are quite similar to those given in §8 for the case of unbounded functions.

Since $L(I; b, c)$ is defined in (30) by different formulas according to which of the cases (28) holds, we must now consider in what intervals $b$ must lie in order that each of these cases may apply, when $I$ and $c$ are given. It is clear from (22) that all four functions

$$p(b; b, c), \qquad p(c; b, c), \qquad q(c; b, c), \qquad q(b; b, c)$$

are increasing functions of $b$. As $b$ increases from $c$ to 1, the four functions increase from 0 to certain limiting values.

$$\begin{aligned} &p(1; 1, c) = \log \frac{(1+c)^2}{4c}, \qquad p(c; 1, c) = \frac{1-c}{1+c} \log \frac{1}{c}, \\ &q(c; 1, c) = \frac{1+c}{1-c} \log \frac{1}{c}, \qquad q(1; 1, c) = +\infty. \end{aligned} \tag{60}$$

We wish to determine values of $b_p, \bar{b}_p, \bar{b}_q, b_q$, not greater than 1, which satisfy

$$p(b_p; b_p, c) = p(c; \bar{b}_p, c) = q(c; \bar{b}_q, c) = q(b_q; b_q, c) = I, \tag{61}$$

so far as this is possible. We can always find $b_q < 1$; and we can find

$$\begin{aligned} &b_p \leqq 1 \quad \text{if} \quad a \geqq 4c/(1+c)^2, \qquad \bar{b}_p \leqq 1 \quad \text{if} \quad a \geqq c^{(1-c)/(1+c)}, \\ &\bar{b}_q \leqq 1 \quad \text{if} \quad a \geqq c^{(1+c)/(1-c)}. \end{aligned} \tag{62}$$

The quantities $b_p, \bar{b}_p, \bar{b}_q, b_q$, so far as they exist, are seen from (23) to satisfy the inequalities

$$b_p > \bar{b}_p > \bar{b}_q > b_q. \tag{63}$$

If any one of the quantities does not exist, we shall treat it in inequalities as if it were more than 1. For example, $\bar{b}_p > 1$ would mean that no $\bar{b}_p \leqq 1$ can be found. If $\bar{b}_p \geqq 1$, then $b \geqq \bar{b}_p$ is impossible, since $b < 1$ in any case; and $b < \bar{b}_p$ would impose no condition on $b$. With this interpretation, the three cases (28) take the form

$$\begin{aligned} &\text{Case } p. \quad b_p \geqq b \geqq \bar{b}_p. \\ (64) \qquad &\text{Case } o. \quad \bar{b}_p \geqq b \geqq \bar{b}_q. \\ &\text{Case } q. \quad \bar{b}_q \geqq b \geqq b_q. \end{aligned}$$

For some values of $a$ and $c$, not all the cases occur.

We now calculate $L_b(I; b, c)$ in each of the three cases. Besides (48) we have

$$(65) \qquad p_b(r; b, c) = \frac{1 - r}{b(1 + r)}, \qquad q_b(r; b, c) = \frac{1 + r}{b(1 - r)}.$$

From these we find that

$$(66) \qquad L_b(I; b, c) = \begin{cases} \dfrac{2(1 + r)}{b(1 - r)} \text{ where } p(r; b, c) = I & (\text{Case } p), \\ \dfrac{2}{bI} \log \dfrac{b}{c} & (\text{Case } o), \\ \dfrac{2(1 - r)}{b(1 + r)} \text{ where } q(r; b, c) = I & (\text{Case } q). \end{cases}$$

We see that at $\bar{b}_p$ and $\bar{b}_q$ the derivative has the same value to the left and to the right.

It is clear that $L_b(I; b, c) > 0$ in all cases, so that $L(I; b, c)$ is a monotone increasing function of $b$. Hence by (29), the smallest value of $J$ for given $a$ and $c$ is obtained by taking $b$ as small as possible. We thus verify that the hyperbolic expansion is maximized in this way.

We next consider the behavior of $L(I; b, c)$ as $b \to 1$. In order for $b \to 1$ to be possible, we must have $b_p \geqq 1$. It may be shown that $L(I; b, c)$ approaches a finite limit if $b_p > 1$, and that $L(I; b, c) + 2 \log (1 - b)$ approaches a finite limit if $b_p = 1$. From this we find that

$$(67) \qquad \begin{aligned} M(I; b, c) &\to + \infty \qquad \text{as } b \to 1 \text{ if } b_p > 1, \\ M(I; b, c) &\to - \infty \qquad \text{as } b \to 1 \text{ if } b_p = 1. \end{aligned}$$

The first formula shows that no upper bound for $d$ in terms of $a$ and $c$ can be found if $b_p > 1$. On the other hand, if $b_p = 1$, then $d \to 0$ as $b \to 1$. In fact, if $d_{\min}$ and $d_{\max}$ denote the smallest and largest values of $d$ for given $a$, $b$, $c$, it may be shown that if $a$ and $c$ have fixed values such that $b_p = 1$, then

$$(68) \qquad d_{\max}/d_{\min} \to 4/c \qquad \text{as } b \to 1.$$

We now turn to the consideration of

$$(69) \qquad M_b(I; b, c) = \frac{2b}{1 - b^2} - L_b(I; b, c).$$

From (66) we see that the condition $M_b(I; b, c) < 0$ at the points $b_p, \bar{b}_p, \bar{b}_q, b_q$ reduces to

$$(70) \qquad b_p < 1, \quad \bar{b}_p < ((1 + c)/2)^{1/2}, \quad \bar{b}_q < ((1 - c)/2)^{1/2}, \quad b_q < 1/2,$$

respectively. Since $p(b_p;\ b_p,\ c) = I$, the first condition is equivalent to $p(1; 1, c) > I$, and similarly for the others. The four conditions become

$$(71) \qquad \begin{aligned} &a > 4c/(1 + c)^2, \qquad a > (2c^2/(1 + c))^{(1-c)/2(1+c)}, \\ &a > (2c^2/(1 - c))^{(1+c)/2(1-c)}, \qquad a > c/2(1 - c)^2. \end{aligned}$$

These conditions are satisfied above the curves which are denoted by $p$, $\bar{p}$, $\bar{q}$, $q$, respectively, in Figure 2. We wish to know that the curves have the relative position shown in the figure. To verify that curve $\bar{p}$ lies below curve $p$, we put

$$\phi(c) = (3 + 5c) \log 2 + 4c \log c - (3 + 5c) \log (1 + c),$$

and have to show $\phi(c) > 0$ for $0 < c < 1$. Calculating the first and second derivatives, we find that $\phi''(c) > 0$ and $\phi'(1) = 0$, hence $\phi'(c) < 0$ for $0 < c < 1$; then since $\phi(1) = 0$, we find that $\phi(c) > 0$ in the interval. By means of similar considerations, we can show that each other pair of curves intersects in exactly one point, and then by numerical calculation it is easily seen that the points of intersection lie as shown in the figure.

By a detailed study of the three cases (64), it may be shown that $M_b(I; b, c) < 0$, if at all, in a single subinterval of the whole interval $b_p \geqq b \geqq b_q$. It is clear then from (70) and (67) that a necessary and sufficient condition that $M(I; b, c)$ should be monotone decreasing is that $b_p \leqq 1$ and $b_q \leqq 1/2$. It is seen that this is also the condition that the largest value of $M(I; b, c)$ is for $b = b_q$.

It may be shown further that $M(I; b, c)$ is decreasing in some subinterval for $(c, a)$ above the heavy broken curve in Figure 2. This curve is tangent to $\bar{q}$. The equation of the curve to the left of the point of tangency is

$$(72) \qquad a = \frac{c}{K(1 - c)^2},$$

where $K = 2.31 \cdots$ is a constant. This part of the curve together with $q$ and $\bar{q}$ bound a region where $M(I; b, c)$ is decreasing somewhere between $b_q$ and $\bar{b}_q$. The equation of the curve to the right of the point of tangency is

$$(73) \qquad c^2 = [1 + I - (2I + I^2)^{1/2}]e^{I-(2I+I^2)^{1/2}},$$

where as usual $I = \log 1/a$. This curve with $\bar{q}$ and $\bar{p}$ bounds a region where $M(I; b, c)$ is decreasing somewhere between $\bar{b}_q$ and $\bar{b}_p$.

The three heavy curves in Figure 2 divide the unit square into 5 regions which are numbered from 1 to 5. In these, $M(I; b, c)$ has the following behaviour: 1, decreasing throughout; 2, increasing and then decreasing; 3, de-

creasing and then increasing; 4, increasing, then decreasing, then increasing again; 5, increasing throughout.

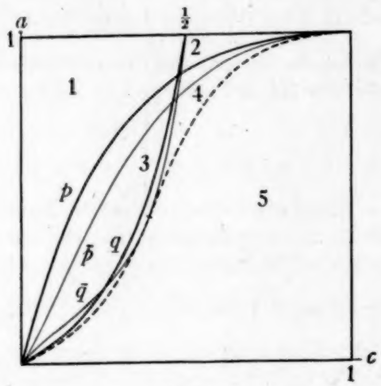


FIG. 2

We come finally to expressing the upper bound for $d$ in terms of $a$ and $c$. Only if $b_p \leqq 1$ is there any upper bound for $d$. In region 1, for $b_p \leqq 1$ and $b_q \leqq 1/2$, the largest $d$ is attained for the smallest possible $b$, as was mentioned in (41). The result is then (38) with $b$ determined from (39). Region 2 is divided into two parts by $\bar{q}$. In the small region to the left we have

$$\log d \leqq \log \frac{1 - c^2}{1 - b^2} - p(1 - 2b^2; b, c), \tag{74}$$

where $b$ is the smaller root of $q(1-2b^2; b, c) = I$. To the right we have

$$\log d \leqq \log \frac{1 - c^2}{1 - b^2} - \frac{b^2}{1 - b^2} \log \frac{b}{c}, \tag{75}$$

where $b$ is the smaller root of $\log b/c = b^2 I/(1-b^2)$.

Finally, we restate the most striking result: *If values of $a$ and $c$ are given for which $b$ cannot have values arbitrarily near to* 1, *then there is a maximum possible value for $d$; and the same is true for values of $a$ and $c$ which are limits of such values. But for all other values of $a$ and $c$, $d$ may have arbitrarily large values.* The surprising part is that there is a sudden jump from one case to the other, rather than a gradual transition.

8. **Appendix.** As a supplement to our study of bounded univalent functions, we now consider univalent functions which are not supposed bounded. The results of this section are obtained from those of §5 by a suitable passage to a limit; no use is made of §§6 and 7.

Let $F(z)$ be a function which is regular and univalent for $|z| < 1$, and for

which $F(0)=0$ and $F'(0)=1$. For any fixed $z_0\neq 0$ in the unit circle, we put

$$[1] \qquad b = |z_0|, \qquad C = |F(z_0)|, \qquad D = |F'(z_0)|.$$

We shall study the relations between $b$, $C$, $D$. Individually, the quantities satisfy only the inequalities

$$[2] \qquad 0 < b < 1, \qquad 0 < C, \qquad 0 < D.$$

Now any such $F(z)$ can be approximated by bounded functions of the same type, if the bound is allowed to vary. Hence any possible values of $b$, $C$, $D$ can be approximated for bounded functions. Conversely, on the basis of [5], the univalent functions form a normal family, so that any values of $b$, $C$, $D$ with $0<b<1$ which can be approximated can also be assumed. Hence the values possible for $b$, $C$, $D$ are those attained for bounded $F(z)$ together with the limit points satisfying $0<b<1$.

If $F(z)$ is bounded, we can choose $a>0$ so that $|aF(z)|<1$ for $|z|<1$. We then put

$$[3] \qquad f(z) = aF(z),$$

so that $f(z)$ is a function of the class previously considered, and $f'(0)=a$. Thus $a$ and $b$ have the same meaning as before, and

$$[4] \qquad c = aC, \qquad d = aD.$$

From the known relations between $a$, $b$, $c$, $d$, we obtain the relations between $a$, $b$, $C$, $D$; by eliminating $a$, we find the relations between $b$, $C$, $D$.

**Relations between $b$ and $C$.** If in formula (31) we put $c=aC$, and then let $a\to 0$, we obtain the well known inequalities[6]

$$[5] \qquad \frac{b}{(1+b)^2} \leqq C \leqq \frac{b}{(1-b)^2}.$$

The bounds are attained for the function

$$[6] \qquad F(z) = \frac{z}{(1-z)^2}$$

for $z_0<0$ and $z_0>0$, respectively. The function [6] maps $|z|<1$ onto the $w$-plane excluding those points for which $w\leqq -1/4$.

**Relations between $b$, $C$, $D$.** We remark first that the required inequalities are not obtained from the relations between $b$, $c$, $d$ by passing to a limit. If in (34) we put $c=aC$ and $d=aD$, and let $a\to 0$, we obtain Nevanlinna's result[7]

---

[6] See for example L. Bieberbach, *Lehrbuch der Funktionentheorie*, vol. 2, chap. 1, §9.

[7] R. Nevanlinna, *Über die konforme Abbildungen von Sterngebieten*, Finska Vetenskaps-Societeten Förhandlingar, vol. 63 (1921), no. 6, p. 18.

$$\text{[7]}\qquad \frac{C(1-b)}{b(1+b)} \leqq D \leqq \frac{C(1+b)}{b(1-b)}.$$

This inequality gives the best bounds for $D/C$ in terms of $b$. It does not however give the best bounds for $D$ in terms of $b$ and $C$, since either of the equalities in (34) is attained only for a certain positive value of $a$.

The sharp lower bound for $D$ in terms of $b$ and $C$ is easily obtained. In (42) we put $c=aC$, $d=aD$, and let $a\to 0$, which gives

$$\text{[8]}\qquad D \geqq \frac{1-b^2}{b^2}C^2.$$

It may be verified that the equality is attained for a function mapping $|z|<1$ on the $w$-plane with slits along the positive and negative real axes. For a given $b$, the slits may be so chosen that $C$ has any possible value. The equality is attained for any positive or negative $z_0$.

We turn now to the problem of finding the upper bound for $D$ in terms of $b$ and $C$. We start by introducing the functions

$$\text{[9]}\qquad \begin{aligned} P(r;b,C) &= \log\frac{r}{C(1+r)^2} + \frac{1-r}{1+r}\log\frac{b}{r}, \\ Q(r;b,C) &= \log\frac{r}{C(1-r)^2} + \frac{1+r}{1-r}\log\frac{b}{r}, \end{aligned}$$

for $0<r\leqq b$; for $r=0$ we put

$$\text{[10]}\qquad P(0;b,C) = Q(0;b,C) = \log b/C,$$

which is the limiting form of [9]. The functions (21) are related to these by the equations

$$\text{[11]}\qquad \begin{aligned} p(r;b,c) &= p(r;b,aC) = P(r;b,C) - \log\frac{a}{(1+aC)^2}, \\ q(r;b,c) &= q(r;b,aC) = Q(r;b,C) - \log\frac{a}{(1-aC)^2}. \end{aligned}$$

The function $P(r;b,C)$ is a decreasing and $Q(r;b,C)$ an increasing function of $r$; both increase with $b$, and decrease as $C$ increases. The partial derivatives have the values

$$\text{[12]}\qquad \begin{aligned} P_r(r;b,C) &= \frac{-2}{(1+r)^2}\log\frac{b}{r}, \quad P_b(r;b,C) = \frac{1-r}{b(1+r)}, \quad P_C(r;b,C) = -\frac{1}{C}. \\ Q_r(r;b,C) &= \frac{2}{(1-r)^2}\log\frac{b}{r}, \quad Q_b(r;b,C) = \frac{1+r}{b(1-r)}, \quad Q_C(r;b,C) = -\frac{1}{C}. \end{aligned}$$

The three cases (28) take the following limiting forms as $a \to 0$:

$$[13] \quad P(b; b, C) \leqq 0 \leqq P(0; b, C), \qquad P(0; b, C) \leqq 0 \leqq Q(0; b, C), \\ Q(0; b, C) \leqq 0 \leqq Q(b; b, C).$$

These are equivalent to

$$[14] \qquad C \leqq b, \qquad C = b, \qquad C \geqq b,$$

respectively, where $C$ of course satisfies [5].

In Cases $p$ and $q$, we determined $r$ from $p(r; b, c) = I$ and $q(r; b, c) = I$, respectively. These equations are the same as

$$[15] \quad P(r; b, C) + 2 \log (1 + aC) = 0, \quad Q(r; b, C) + 2 \log (1 - aC) = 0.$$

As $a \to 0$, their roots approach those of

$$[16] \qquad P(r; b, C) = 0, \qquad Q(r; b, C) = 0.$$

Putting $c = aC$ and $d = aD$ in (43), and letting $a \to 0$, we find that

$$[17] \qquad \log D \leqq M(b, C)$$

where

$$[18] \qquad M(b, C) = \log \frac{1}{1 - b^2} - L(b, C),$$

$L(b, C)$ being defined for all possible values of $b$ and $C$ by

$$[19] \quad L(b, C) = \begin{cases} Q(r; b, C) \text{ where } r \text{ satisfies } P(r; b, C) = 0, \text{ if } C \leqq b, \\ P(r; b, C) \text{ where } r \text{ satisfies } Q(r; b, C) = 0, \text{ if } C \geqq b. \end{cases}$$

These two formulas correspond to Cases $p$ and $q$ in (30). For the case $C = b$, either of these, and also Case $o$, leads to the result

$$[20] \qquad D \leqq \frac{1}{1 - b^2} \qquad \text{if } C = b.$$

It may be verified that the equality in [20] is attained for a function $F(z)$ which maps $|z| < 1$ on the $w$-plane slit to infinity at one or both ends of the perpendicular bisector of the segment joining 0 and $f(z_0)$. We see from [8] and [17] that when $C$ has its smallest value, we must have $D = (1-b)/(1+b)^3$, and that when $C$ has its largest value, $D = (1+b)/(1-b)^3$; in these cases the equalities in both [8] and [17] are attained for the function [6]. We know that in any case, there is some function for which the equality in [17] is attained; but the extremal function does not seem to be of a very simple sort except in the three cases mentioned.

**Relations between $b$ and $D$.** From [8] we see that $D$ has its smallest value when $C$ has its smallest value, determined from [5]. This gives

$$D \geqq \frac{1-b}{(1+b)^3}. \tag{21}$$

The equality is attained for the function [6] with $z_0<0$.

We next determine the upper bound for $D$. In the first part of [19], as $C$ increases, $r$ decreases, and hence $Q(r; b, C)$ decreases; in the second part, as $C$ increases, $r$ increases, and hence $P(r; b, C)$ decreases. Hence in either case, $L(b, C)$ is decreasing, or $M(b, C)$ is increasing. Thus the largest value of $D$ is obtained when $C$ has its largest value, and hence

$$D \leqq \frac{1+b}{(1-b)^3}. \tag{22}$$

The equality is attained for the function [6] with $z_0>0$.

We may also verify that $M(b, C)$ is increasing by calculating its partial derivative. Making use of [12], we find that

$$M_C(b, C) = \begin{cases} \dfrac{1}{C}\left[\left(\dfrac{1+r}{1-r}\right)^2+1\right] & \text{where } P(r; b, C) = 0, \text{ if } C \leqq b, \\ \dfrac{1}{C}\left[\left(\dfrac{1-r}{1+r}\right)^2+1\right] & \text{where } Q(r; b, C) = 0, \text{ if } C \geqq b, \end{cases} \tag{23}$$

so that $M_C(b, C)>0$ throughout.

The inequalities [21] and [22] ("distortion theorem") are well known, and were used to derive [5] in the original approach to this subject(8).

**Remarks about $D/C$ and $D/C^2$.** From [8] we see that $D/C$ has its smallest value when $C$ is smallest. From [23] we see that $M(b, C)-\log C$ is an increasing function, and hence $D/C$ has its largest value when $C$ is largest. Thus we are again led to [7].

On the other hand, [8] shows that $D/C^2$ can reach its smallest value for any $C$. From [23] we see that $M(b, C)-2\log C$ has its maximum for $C=b$, so that $D/C^2$ attains its largest value only in this case. We obtain the inequalities

$$\frac{1-b^2}{b^2} \leqq \frac{D}{C^2} \leqq \frac{1}{b^2(1-b^2)}. \tag{24}$$

This result gives the bounds for the derivative of $1/F(z)$, or for the derivative of a function univalent in the exterior of the unit circle and leaving $\infty$ fixed. The problem was solved in this form by Löwner(9) (without using the "method of Löwner").

---

(8) See Bieberbach, loc. cit.

(9) K. Löwner, *Über Extremumsätze bei der konformen Abbildungen des Äusseren des Einheitskreises*, Mathematische Zeitschrift, vol. 3 (1919), pp. 65–77.

**Relations between** $C$ **and** $D$. To find the lower bound for $D$ in terms of $C$, we must eliminate $b$ from [8]. Evidently the larger $b$ is, the smaller $D$ may be. Now $b$ is restricted only by [5]. If $C \geqq 1/4$, $b$ may have values arbitrarily near to 1, and hence there is no positive lower bound for $D$. But if $C<1/4$, then the lower bound for $D$ is given by [8] with $b$ determined from $b/(1+b)^2 = C$. Solving for $b$ and substituting gives

[25] $$2D \geqq 1 - 4C + (1 - 2C)(1 - 4C)^{1/2} \qquad \text{for } C < 1/4.$$

The equality is attained for the function [6] with $z_0<0$.

We now turn to the problem of finding the upper bound for $D$ in terms of $C$. We first consider the behavior of $L(b, C)$ as $b \to 1$. In order that $b \to 1$ should be possible, we must have $C \geqq 1/4$. If $C \geqq 1$, then the second part of [19] applies. As $b \to 1$, $r$ decreases, and $L(b, C)$ increases to a finite limit. If $C<1$, then the first part of [19] applies for $b$ near 1. As $b \to 1$, $r$ increases, and $L(b, C)$ increases; and $L(b, C)$ approaches a finite limit unless $r \to 1$. Now if $r \to 1$ as $b \to 1$, we must have $P(1; 1, C)=0$, or $C=1/4$. Hence if $C>1/4$, $L(b, C)$ increases to a finite limit as $b \to 1$. The case $C=1/4$ remains to be considered. Here $r$ is determined from $P(r; b, 1/4)=0$, which is equivalent to

[26] $$\log b = \log r + \frac{1+r}{1-r} \log \frac{(1+r)^2}{4r} \cdot$$

From this we find that

[27] $$1 - r \sim 2(1 - b) \qquad \text{as } b \to 1.$$

Using this in the formula $L(b, 1/4)=Q(r; b, 1/4)$, we find that

[28] $$L(b, 1/4) + 2 \log (1 - b) \to 1 \qquad \text{as } b \to 1.$$

From these results we find that

[29] $$\begin{aligned} M(b, C) &\to + \infty \qquad \text{as } b \to 1 \text{ if } C > 1/4, \\ M(b, C) &\to - \infty \qquad \text{as } b \to 1 \text{ if } C = 1/4. \end{aligned}$$

The first formula shows that no upper bound for $D$ can be found if $C>1/4$. The second-part may be written more accurately as

[30] $$M(b, 1/4) - \log (1 - b) \to - 1 - \log 2 \qquad \text{as } b \to 1.$$

If we denote by $D_{\min}$ and $D_{\max}$ the smallest and largest values of $D$ which are possible for given $b$ and $C$, then for $C=1/4$ we have from [8] and [30]

[31] $$D_{\min} \sim \frac{1-b}{8}, \qquad D_{\max} \sim \frac{1-b}{2e}, \qquad \frac{D_{\max}}{D_{\min}} \to \frac{4}{e} \qquad \text{as } b \to 1.$$

We now consider the derivative of $M(b, C)$. Using [12], we find that

$$[32]\qquad M_b(b,C)=\begin{cases}\dfrac{2b}{1-b^2}-\dfrac{2(1-r)}{b(1+r)} & \text{where } Q(r;b,C)=0, \text{ if } b\leqq C,\\[2ex] \dfrac{2b}{1-b^2}-\dfrac{2(1+r)}{b(1-r)} & \text{where } P(r;b,C)=0, \text{ if } b\geqq C.\end{cases}$$

Hence at $b=C$, $M_b(b, C)$ has the same value to the left and to the right.

A remark which will be useful below is the following. For any fixed $b$ the root $r$ of $P(r; b, C)=0$ decreases from $b$ to 0 as $C$ increases from its smallest possible value to $b$; and the root $r$ of $Q(r; b, C)=0$ increases from 0 to $b$ as $C$ increases from $b$ to its largest possible value. Hence if $0\leqq r\leqq b$, the equation $P(r; b, C)=0$ determines a value of $C\leqq b$, and $Q(r; b, C)$ determines a $C\geqq b$, both satisfying [5].

Now the equation

$$M_b(b, C) = 0$$

requires in the first case that $r=1-2b^2$ and in the second that $r=2b^2-1$. Since we must have $0\leqq r\leqq b$, $M_b(b, C)$ vanishes only in the following cases:

$$[33]\qquad M_b(b,C)=0\begin{cases}\text{if } 1/2\leqq b\leqq 2^{-1/2} & \text{and } Q(1-2b^2;b,C)=0,\\ \text{if } 2^{-1/2}\leqq b<1 & \text{and } P(2b^2-1;b,C)=0.\end{cases}$$

These may be combined in the statement that $M_b(b, C)=0$ only along the curve

$$[34]\qquad 4C = b^{-5+1/b^2}\,|\,1-2b^2\,|^{2-1/b^2},\qquad 1/2\leqq b<1,$$

which from the preceding paragraph must lie between the bounds [5]. To study this curve, we put $h=1/b^2$ and

$$\phi(h) = 2\log 4C.$$

Then

$$\phi(h) = (h+1)\log h - 2(h-2)\log|h-2|,$$

where for $h=2$ we interpret the right side as its limiting value $3\log 2$. Differentiating, we have

$$\phi'(h) = \log h - 2\log|h-2| - 1 + \frac{1}{h},\qquad \phi''(h) = \frac{1}{h} - \frac{2}{h-2} - \frac{1}{h^2}\cdot$$

We see that $\phi''(h)>0$ for $1<h<2$, and $\phi''(h)<0$ for $2<h<4$. Using this we find that $\phi'(h)$ increases from 0 to $+\infty$ as $h$ increases from 1 to 2, and then decreases to $-3/4$ as $h$ increases to 4. We must have $\phi'(h)=0$ at some point between 2 and 4, and in fact for $h=3.27\cdots$; and here $\phi(h)$ has its maximum. Therefore, in [34], $C$ increases from 2 when $b=1/2$ to a maximum $K=2.31\cdots$ for $b=0.55\cdots$, and then decreases to 1/4 as $b$ increases to 1. Figure 3 shows

[34] as a broken curve and the bounds [5] as solid curves.

The curve [34] divides the region defined by [5] into two parts. There is no difficulty in seeing that $M_b(b, C) < 0$ in the lower part, and $M_b(b, C) > 0$ in the upper part. Hence $M(b, C)$ is a decreasing function of $b$ if $C \leqq 1/4$; if $1/4 < C \leqq 2$, it first decreases, and then increases to $+\infty$; if $2 < C < K$, it first

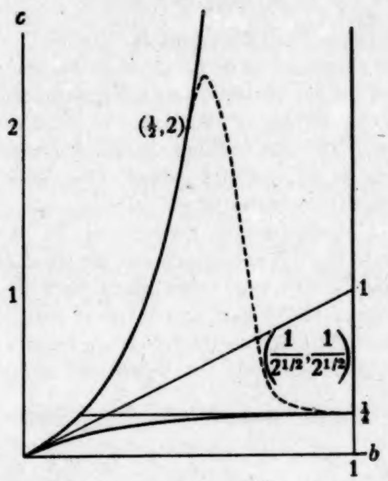


FIG. 3

increases, then decreases, then increases to $+\infty$; and if $C \geqq K$, $M(b, C)$ increases throughout, and approaches $+\infty$ as $b \to 1$. There is an upper bound for $D$ only if $C \leqq 1/4$, and then it is attained when $b$ has its smallest possible value. Hence the bound is given by [22] with $b$ determined from $b/(1-b)^2 = C$. Substituting this value of $b$, we find that

$$[35] \qquad 2D \leqq 1 + 4C + (1 + 2C)(1 + 4C)^{1/2} \qquad \text{for } C \leqq 1/4.$$

The equality is attained for the function [6] with $z_0 > 0$. It is to be noted that the upper bound for $D$ in terms of $C$ increases from 1 to $2.06 \cdots$ as $C$ increases from 0 to 1/4, and then jumps to $+\infty$.

UNIVERSITY OF CALIFORNIA,
BERKELEY, CALIF.

# ON THE OSCILLATION OF DIFFERENTIAL TRANSFORMS. I

BY

G. SZEGÖ

## 1. Introduction

1.1. In a recent paper G. Pólya and N. Wiener[1] studied the relation between the analytic character of a real periodic function and the number of the sign variations of its derivatives. The purpose of the present paper is to develop another way of attacking this problem different from that used by the authors mentioned. It leads to a new proof of Theorem 1 of the paper of Pólya and Wiener and to refinements of their Theorems 2 and 3 which are in a certain sense best possible results[2].

Let $f(x)$ be a real periodic function with period $2\pi$ for which all derivatives $f^{(k)}(x)$ exist. We denote by $2N_k$ the number of the mod $2\pi$ distinct values of $x$ for which a sign variation of $f^{(k)}(x)$ takes place. In what follows we give first a new proof of Theorem 1 of Pólya and Wiener. A further, more elaborate, application of our method leads to the following results which correspond to the Theorems 3 and 2, respectively, of the authors mentioned.

Theorem A. *Let $N_k < k/\log k$ provided $k$ is sufficiently large. Then $f(x)$ is an integral function.*

Theorem B. *Let $\rho > 1$ and let $N_k < (k/\rho)^{1/\rho}/2$ provided $k$ is sufficiently large. Then $f(x)$ is an integral function of order not greater than $\rho/(\rho-1)$.*

The following results are more informative.

Theorem A′. *Let for sufficiently large $k$*

$$N_k < \frac{k}{\log k}\left(1 + \frac{\log\log k - \omega(k)}{\log k}\right) \tag{1.1.1}$$

*where $\omega(k) \to +\infty$. Then the conclusion of Theorem* A *holds.*

Theorem B′. *Let $\rho > 1$ and let $p$ be a positive number such that $p\rho^{2+1/\rho} > 1$. If for sufficiently large $k$*

$$N_k < (k/\rho)^{1/\rho}\left(1 - p\,\frac{\log k}{k^{1-1/\rho}}\right), \tag{1.1.2}$$

*then the conclusion of Theorem* B *holds.*

Presented to the Society, November 22, 1941; received by the editors January 2, 1942.

[1] G. Pólya and N. Wiener, *On the oscillation of the derivatives of a periodic function*, these Transactions, vol. 52 (1942), pp. 249–256.

[2] See the counterexamples given below, section 7.

The following result contains Theorem A′ (therefore also Theorem A):

THEOREM A″. *Let $H$ be a constant such that*

$$H + \log((1/2)\log 2) > 0, \tag{1.1.3}$$

*and let for sufficiently large $k$*

$$N_k < \frac{k}{\log k}\left(1 + \frac{\log\log k - H}{\log k}\right). \tag{1.1.4}$$

*Then the function $f(x)$ is analytic in the strip*

$$|\Im f(x)| < H + \log((1/2)\log 2). \tag{1.1.5}$$

1.2. In various conversations Professor Hille suggested certain analogues of these problems considering $\vartheta^l f(x)$ instead of $f^{(2l)}(x)$ where $\vartheta$ is a given second order differential operator satisfying suitable conditions([3]). In the last part of the present paper we illustrate the further applicability of our method by discussing the special operator

$$\vartheta = (1 - x^2)D^2 - 2xD, \qquad D = d/dx, \tag{1.2.1}$$

the "characteristic functions" of which are the classical Legendre functions. Let $f(x)$ be a function having derivatives of all orders in $-1 \leqq x \leqq +1$ and let $N_k = N_{2l}$ denote the number of the sign variations of $\vartheta^l f(x)$ in the interval $-1 < x < +1$. Then we prove([4])

THEOREM C. *If $N_k \leqq N$, $k$ sufficiently large and even, then $f(x)$ must be a polynomial of degree less than or equal to $N$.*

THEOREM D. *If $N_k$ satisfies the condition of Theorem* A″, *$k$ even, then $f(x)$ is analytic in an ellipse with foci at $-1$ and $+1$ the sum of the semi-axes of which is*

$$\exp\{H + \log((1/2)\log 2)\}.$$

These results correspond to Theorem 1 of Pólya and Wiener and to Theorem A″, respectively. The analogue of Theorem B can also be dealt with.

1.3. In what follows we give the proofs of the results formulated above. Section 2 contains a new proof of Theorem 1 of Pólya and Wiener; the underlying idea of this proof is used throughout the present paper. Section 3 contains the proof of Theorem A″, section 4 that of Theorem B′. Sections 5 and 6 are devoted to the proofs of Theorems C and D involving Legendre's operator. Finally in section 7 certain counterexamples are exhibited which

---

([3]) See below, pp. 463–497.

([4]) The proof furnishes the conclusion of Theorem C under the condition that $N_k \leqq N$ holds for an infinite number of $k$ values. (The same holds in section 2.)

show that the conditions of Theorems A and B on $N_k$ cannot be replaced by $N_k=O(k)$ and $N_k=O(k^\alpha)$, $\alpha>1/\rho$, respectively.

## 2. New proof of Theorem 1 of Pólya and Wiener

2.1. Let

$$f(x) = \sum_{\nu=-\infty}^{+\infty} c_\nu e^{i\nu x}, \qquad c_{-\nu} = \bar{c}_\nu, \tag{2.1.1}$$

and let $2N_k$ be the number of the mod $2\pi$ distinct sign variations of $f^{(k)}(x)$. We assume that $k$ goes to $+\infty$ through a sequence of integers such that $N_k$ has a constant value $N$. Then we show that $f(x)$ is a trigonometric polynomial of degree less than or equal to $N$, that is, we prove $c_{N+m}=0$, $m>0$.

Let $x_1, x_2, \cdots, x_{2N}$ denote the mod $2\pi$ distinct sign variations of $f^{(k)}(x)$, that is, the values of $x$ for which $f^{(k)}(x)$ changes its sign; $x_\nu=x_\nu(k)$. Let $\alpha$ be real and (5)

$$\begin{aligned} u(x) &= \sin\frac{x-x_1}{2}\sin\frac{x-x_2}{2}\cdots\sin\frac{x-x_{2N}}{2}(1+\cos^m(x+\alpha)) \\ &= \sum_{\nu=-N-m}^{N+m} u_\nu e^{i\nu x}, \qquad u_{-\nu} = \bar{u}_\nu. \end{aligned} \tag{2.1.2}$$

(In case $N=0$ we write $u(x)=1+\cos^m(x+\alpha)$.) This is a trigonometric polynomial of the fixed degree $N+m$, the sign variations of which are the same as those of $f^{(k)}(x)$. The coefficients $u_\nu=u_\nu(k)$ are bounded as $k\to\infty$; this can easily be showed by multiplying out the expression

$$u(x)=2^{-2N}\prod_{\nu=1}^{2N}\left(e^{-i(\pi+x_\nu)/2}e^{ix/2}+e^{i(\pi+x_\nu)/2}e^{-ix/2}\right)\left(1+2^{-m}(e^{i\alpha}e^{ix}+e^{-i\alpha}e^{-ix})^m\right). \tag{2.1.3}$$

Also we obtain for the highest coefficient of $u(x)$

$$\begin{aligned} u_{N+m} &= (-1)^N 2^{-2N-m}\exp\left\{-i\sum x_\nu/2+im\alpha\right\} \\ &= (-1)^N 2^{-2N-m}\exp\left\{ix_0+im\alpha\right\} \end{aligned} \tag{2.1.4}$$

where the real quantity $x_0=x_0(k)$ depends on $k$ but it is independent of $\alpha$.

2.2. Let $c_{N+m}\neq 0$. The sign of

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi} f^{(k)}(x)u(x)dx = \sum_{\nu=-N-m}^{N+m}(i\nu)^k c_\nu u_{-\nu} \tag{2.2.1}$$

is *independent* of $\alpha$, positive say. We determine $\alpha$ in such a way that the last term

(5) We could use as well $1+\cos m(x+\alpha)$ or $(1+\cos(x+\alpha))^m$ instead of $1+\cos^m(x+\alpha)$.

$$(2.2.2) \qquad \begin{aligned}&(i(N+m))^k c_{N+m} u_{-N-m}\\ &\qquad = (i(N+m))^k c_{N+m}(-1)^N 2^{-2N-m} \exp\{-ix_0 - im\alpha\}\end{aligned}$$

becomes *real and negative*. Then

$$(2.2.3) \qquad 2(N+m)^k |c_{N+m}| 2^{-2N-m} < \sum_{\nu=-N-m+1}^{N+m-1} |\nu|^k |c_\nu| |u_{-\nu}| = \sum_{\nu=1}^{N+m-1} \nu^k O(1)$$

follows where the bounds $O(1)$ are independent of $k$. But this involves a contradiction for sufficiently large $k$.

## 3. Proof of Theorem A''

3.1. We start with some preliminary remarks.

(a) The constant $H$ must be positive since $\log((1/2)\log 2) < 0$.

(b) Let $x = \sigma + it$, $\sigma$ and $t$ real, and let $T$ denote the unique value such that (2.1.1) converges for $|t| < T$ and diverges for $|t| > T$ (or $T$ is the largest value such that $f(x)$ is analytic in the strip $|t| < T$). We have

$$(3.1.1) \qquad \limsup_{\nu\to+\infty} |c_\nu|^{1/\nu} = e^{-T}.$$

The modifications necessary for $T=0$ or $T=\infty$ are obvious.

Now another form of the assertion of Theorem A'' is

$$(3.1.2) \qquad T \geqq H + \log((1/2)\log 2).$$

(c) Theorem A' is obviously a consequence of Theorem A''.

3.2. We assume

$$(3.2.1) \qquad \limsup_{\nu\to+\infty} |c_\nu| e^{\nu\gamma} = +\infty, \qquad \gamma > 0,$$

and show that

$$(3.2.2) \qquad \gamma \geqq H + \log((1/2)\log 2).$$

From (3.2.1) we conclude in a well known manner([6]) the existence of a sequence of integers $\{M\}$ such that

$$(3.2.3) \qquad |c_M| e^{M\gamma} > |c_\nu| e^{|\nu|\gamma}, \qquad \pm\nu = 0, 1, 2, \cdots, M-1.$$

Now let $\epsilon$ be an arbitrary but fixed positive number. We define a sequence of integers $k = k(M)$ by

$$(3.2.4) \qquad k = k(M) = [M(\log M + H - \epsilon)].$$

([6]) Cf. G. Pólya and G. Szegö, *Aufgaben und Lehrsätze aus der Analysis*, vol. 1, 1925, p. 18; p. 173, Problem 107. What is needed here is much less than the lemma used by Pólya and Wiener, loc. cit., p. 252.

Then an easy calculation shows that

$$N_k < \frac{k}{\log k}\left(1 + \frac{\log\log k - H}{\log k}\right)$$
$$= M\left(1 + \frac{H - \epsilon - \log\log M}{\log M} + O\left(\left(\frac{\log\log M}{\log M}\right)^2\right)\right)$$
$$\cdot\left(1 + \frac{\log\log M - H}{\log M} + O\left(\left(\frac{\log\log M}{\log M}\right)^2\right)\right)$$
$$(3.2.5)\qquad = M\left(1 - \frac{\epsilon}{\log M} + O\left(\left(\frac{\log\log M}{\log M}\right)^2\right)\right) < M$$

provided $M$ is sufficiently large.

3.3. Let us denote again by $x_1, x_2, \cdots, x_{2N}$, $N=N_k$, the mod $2\pi$ distinct sign variations of $f^{(k)}(x)$, $x_\nu = x_\nu(k)$; here $k=k(M)$. We write, $\alpha$ real(7),

$$(3.3.1)\qquad \begin{aligned} u(x; M) &= \sin\frac{x - x_1}{2}\sin\frac{x - x_2}{2}\cdots\sin\frac{x - x_{2N}}{2}(1 + \cos^m(x + \alpha)) \\ &= \sum_{\nu=-M}^{+M} u_\nu e^{i\nu x}, \qquad u_{-\nu} = \bar{u}_\nu,\ N + m = M. \end{aligned}$$

(For $N=0$ we omit the sine factors.) Since $N<M$, $m$ is positive. The trigonometric polynomial $u(x; M)$ is of degree $M$ and it has the same sign variations as $f^{(k)}(x)$. We prove

LEMMA 1. *Let the trigonometric polynomial $u(x; M)$ be defined by* (3.3.1) *and let*

$$(3.3.2)\qquad U(x; M) = (\cos(x/2))^{2N}(1 + \cos^m x) = \sum_{\nu=-M}^{+M} U_\nu e^{i\nu x} = \sum_{\nu=-M}^{+M} U_\nu \cos\nu x;$$

*then the inequalities*

$$(3.3.3)\qquad |u_\nu| \leqq U_\nu, \qquad \nu = 0, 1, 2, \cdots, M,$$

*hold, with the sign "=" for $\nu=M$.*

Indeed as (2.1.3) shows, the coefficients $u_\nu$ of $u(x; M)$ are multilinear functions of $e^{\pm i(\pi + x_\nu)/2}$ and $e^{\pm i\alpha}$ with non-negative coefficients. Thus we do not decrease $|u_\nu|$ by replacing the quantities $e^{\pm i(\pi + x_\nu)/2}$ and $e^{\pm i\alpha}$ by 1, or by replacing the constants $x_\nu$ by $-\pi$ and $\alpha$ by 0. This leads precisely to (3.3.2). The assertion regarding $|u_M| = U_M$ is also clear. We have (see (2.1.4))

$$(3.3.4)\qquad \begin{aligned} u_M &= (-1)^N 2^{-2N-m}\exp\{ix_0 + im\alpha\}, \qquad x_0 = -\sum x_\nu/2, \\ U_M &= 2^{-2N-m} = 2^{-N-M}. \end{aligned}$$

(7) See Footnote 5.

3.4. Now

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi} f^{(k)}(x)u(x; M)dx = \sum_{\nu=-M}^{+M} (i\nu)^k c_\nu u_{-\nu} \tag{3.4.1}$$

has a sign *independent* of $\alpha$. Choosing $\alpha$ in a proper way ($c_M \neq 0$) and using (3.3.3) we obtain

$$\sum_{\nu=-M+1}^{M-1} |\nu|^k |c_\nu| U_\nu \geqq 2M^k |c_M| U_M. \tag{3.4.2}$$

Taking (3.2.3) and the inequality $x \leqq e^{x-1}$ into account we find

$$\begin{aligned} 2U_M &\leqq \sum_{\nu=-M+1}^{M-1} U_\nu \exp \{(|\nu|/M - 1)k + (M - |\nu|)\gamma\} \\ &= \sum_{\nu=-M+1}^{M-1} U_\nu Q^{|\nu|-M} < Q^{-M} \sum_{\nu=-M+1}^{M-1} U_\nu(Q^\nu + Q^{-\nu}), \qquad Q = e^{k/M-\gamma}. \end{aligned} \tag{3.4.3}$$

From (3.2.4) we conclude that $Q = Q(M) \to \infty$ or more precisely $Q(M) \cong e^{H-\epsilon-\gamma}M$ as $M \to \infty$. (The symbol $a(M) \cong b(M)$ means that $a(M)[b(M)]^{-1} \to 1$ as $M \to \infty$.) Introducing

$$\xi = \xi(M) = (Q + Q^{-1})/2, \qquad T_{|\nu|}(\xi) = (Q^\nu + Q^{-\nu})/2 \tag{3.4.4}$$

where $T_{|\nu|}(\xi)$ is identical with the Tchebichef polynomial, by virtue of (3.3.2), we can write (3.4.3) in the following form:

$$U_M Q^M \leqq \sum_{\nu=-M+1}^{M-1} U_\nu T_{|\nu|}(\xi) = 2^{-N}(1 + \xi)^N(1 + \xi^m) - 2U_M T_M(\xi); \tag{3.4.5}$$

hence

$$2U_M Q^M \leqq 2^{-N}(1 + \xi)^N(1 + \xi^m)$$

or (cf. (3.3.4))

$$2^{1-M}(Q/\xi)^M \leqq (1 + \xi^{-1})^N(1 + \xi^{-m}) < e^{\xi^{-1}M}(1 + \xi^{-1}). \tag{3.4.6}$$

Now let $M \to \infty$. Then

$$2^{-M}(Q/\xi)^M = (1 + Q^{-2})^{-M} \to 1 \tag{3.4.7}$$

since $Q^{-2} = O(M^{-2})$. Further $\xi^{-1}M \to 2e^{-H+\epsilon+\gamma}$ so that

$$2 \leqq \exp\{2e^{-H+\epsilon+\gamma}\} \tag{3.4.8}$$

follows. Since $\epsilon$ is arbitrarily small, this involves (3.2.2).

## 4. Proof of Theorem B′

4.1. Let the order $\lambda$ of $f(x)$ be greater than $\rho/(\rho-1)$. Then using the pre-

vious notation (2.1.1),

$$\liminf_{\nu\to+\infty} \frac{\log\log 1/|c_\nu|}{\log \nu} = \frac{\lambda}{\lambda-1} < \rho \tag{4.1.1}$$

holds(⁹). Consequently

$$\limsup_{\nu\to+\infty} |c_\nu| e^{\nu^\rho} = \infty. \tag{4.1.2}$$

We obtain now instead of (3.2.3) the inequalities

$$|c_M| e^{M^\rho} > |c_\nu| e^{|\nu|^\rho}, \qquad \pm\nu = 0, 1, 2, \cdots, M-1, \tag{4.1.3}$$

holding for a certain sequence $\{M\}$ of integers.

The previous proof needs only unessential modifications.

4.2. We write

$$k = k(M) = \left[\rho M^\rho\left(1 + q\frac{\log M}{M^{\rho-1}}\right)\right] \tag{4.2.1}$$

where $q$ is a fixed constant satisfying the conditions

$$1/\rho < q < p\rho^{1+1/\rho}. \tag{4.2.2}$$

An easy calculation shows that for large $k$

$$\begin{aligned} N_k &< (k/\rho)^{1/\rho}\left(1 - p\frac{\log k}{k^{1-1/\rho}}\right) \\ &= M\left(1 + q\rho^{-1}\frac{\log M}{M^{\rho-1}} + o\left(\frac{\log M}{M^{\rho-1}}\right)\right) \\ &\quad\cdot\left(1 - p\rho^{1/\rho}\frac{\log M}{M^{\rho-1}} + o\left(\frac{\log M}{M^{\rho-1}}\right)\right) \\ &= M\left(1 - (p\rho^{1/\rho} - q\rho^{-1})\frac{\log M}{M^{\rho-1}} + o\left(\frac{\log M}{M^{\rho-1}}\right)\right) < M. \end{aligned} \tag{4.2.3}$$

Using the same notation and the same argument as in §3.4 we obtain instead of (3.4.3)

$$2U_M \leqq \sum_{\nu=-M+1}^{M-1} U_\nu \exp\{(|\nu|/M - 1)k + M^\rho - |\nu|^\rho\}. \tag{4.2.4}$$

Since $M^\rho - |\nu|^\rho \leqq (M - |\nu|)\rho M^{\rho-1}$ we find as before

$$2U_M \leqq R^{-M}\sum_{\nu=-M+1}^{M-1} U_\nu(R^\nu + R^{-\nu}), \qquad R = e^{k/M - \rho M^{\rho-1}}. \tag{4.2.5}$$

(⁹) See Pólya and Wiener, loc. cit., p. 254.

On account of (4.2.1) we have $R=R(M)\cong M^{\rho q}\rightarrow\infty$ as $M\rightarrow\infty$.

Now let

(4.2.6) $$\eta = \eta(M) = (R + R^{-1})/2 \rightarrow \infty \qquad \text{as } M \rightarrow \infty;$$

then we obtain (cf. (3.4.6))

(4.2.7) $$2^{1-M}(R/\eta)^M \leqq e^{\eta^{-1}M}(1 + \eta^{-1}).$$

But $\rho q>1$ so that $(1+R^{-2})^{-M}\rightarrow 1$. Moreover

(4.2.8) $$\eta^{-1}M \cong 2M^{1-\rho q} \rightarrow 0.$$

This furnishes the contradictory inequality $2\leqq 1$.

## 5. Proof of Theorem C

5.1. The proofs of Theorems C and D are based on arguments similar to those followed in the previous part. Instead of trigonometric series, expansions in Legendre series are used.

Let

(5.1.1) $$f(x) = \sum_{\nu=0}^{\infty} c_\nu P_\nu(x),$$

$c_\nu$ real, be the Legendre expansion of $f(x)$ where $P_\nu(x)$ is the Legendre polynomial in the customary notation. By using the notation (1.2.1)

(5.1.2) $$\vartheta^l f(x) = \sum_{\nu=0}^{\infty} (-\lambda_\nu)^l c_\nu P_\nu(x), \qquad \lambda_\nu = \nu(\nu + 1),$$

follows. Let $N_k=N_{2l}$ denote the number of the sign variations of $\vartheta^l f(x)$ in $-1<x<+1$ and let $N_k=N_{2l}=N$ be fixed as $l\rightarrow\infty$ through a proper sequence of integers. We show then that $c_{N+m}=0$, $m>0$.

Let $x_1, x_2, \cdots, x_N$ be the sign variations of $\vartheta^l f(x)$ in $-1<x<+1$. We form(9)

(5.1.3) $$v(x) = (x - x_1)(x - x_2) \cdots (x - x_N)(1 + \delta x^m) = \sum_{\nu=0}^{N+m} v_\nu P_\nu(x)$$

where $\delta$ is either $+1$ or $-1$. This is a polynomial of degree $N+m$ with the same sign variations as $\vartheta^l f(x)$. The coefficients $v_\nu=v_\nu(l)$ are bounded as $l\rightarrow\infty$. Furthermore $v_{N+m}=\delta h_{N+m}$ if $h_\nu$ denotes the highest coefficient in the Legendre expansion of $x^\nu$.

Now

(5.1.4) $$\int_{-1}^{+1} \{\vartheta^l f(x)\} v(x)dx = \sum_{\nu=0}^{N+m} (\nu + 1/2)^{-1}(-\lambda_\nu)^l c_\nu v_\nu.$$

(9) We could use $1+\delta P_m(x)$ instead of $1+\delta x^m$. See Footnote 5.

This expression has the same sign whether $\delta=+1$ or $\delta=-1$. We obtain by a suitable choice of $\delta$

$$(5.1.5)\qquad \sum_{\nu=0}^{N+m-1}(\nu+1/2)^{-1}\lambda_\nu^l|c_\nu||v_\nu| \geqq (N+m+1/2)^{-1}\lambda_{N+m}^l|c_{N+m}|h_{N+m}.$$

Division through by $\lambda_{N+m}^l$ leads to a contradiction as $l\to\infty$ unless $c_{N+m}=0$.

## 6. Proof of Theorem D

6.1. For the proof of Theorem D we follow again the previous argument. Under the assumption (3.2.1) we obtain a sequence $\{M\}$ of integers such that (3.2.3) holds. The definition of $k=2l$ is in this case slightly different from (3.2.4), namely

$$(6.1.1)\qquad k = 2l = 2[(M/2)(\log M + H - \epsilon)].$$

Then $k$ is even and $N=N_k<M$. Now we define $m$ by $N+m=M$ and $v(x)=v(x; M)$ by (5.1.3). We prove

Lemma 2. *Let the rational polynomial* $v(x)=v(x; M)$ *be defined by* (5.1.3) *and let*

$$(6.1.2)\qquad V(x; M) = (x+1)^N(1+x^m) = \sum_{\nu=0}^{M} V_\nu P_\nu(x);$$

*then the inequalities*

$$(6.1.3)\qquad |v_\nu| \leqq V_\nu, \qquad \nu = 0, 1, 2, \cdots, M,$$

*hold with the sign* "$=$" for $\nu=M$.

It is well known([10]) that $P_r(x)P_s(x)$ expanded in terms of Legendre polynomials has non-negative coefficients. Multiplying out

$$(6.1.4)\qquad v(x; M) = \prod_{\nu=1}^{N}(P_1(x)-x_\nu)\{1+\delta(P_1(x))^m\}$$

we see that the coefficients $v_\nu$ of $v(x; M)$ are multilinear functions of $-x_\nu$ and $\delta$ with non-negative coefficients. Obviously we do not decrease $|v_\nu|$ by replacing $-x_\nu$ and $\delta$ by 1 which leads precisely to (6.1.2).

6.2. Starting from (5.1.4) we obtain (5.1.5) and

$$(6.2.1)\qquad (M+1/2)^{-1}h_M \leqq \sum_{\nu=0}^{M-1}(\nu+1/2)^{-1}(\lambda_\nu/\lambda_M)^{k/2}e^{(M-\nu)\gamma}V_\nu.$$

([10]) See J. C. Adams, *On the expression of the product of any two Legendre's coefficients by means of a series of Legendre's coefficients*, Proceedings of the Royal Society, vol. 27 (1878), pp. 63–71; *Collected Scientific Papers*, vol. 1, pp. 487–496. See E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis*, 4th edition, 1935, p. 331, Problem 11.

Now let

$$g_0 = 1; \quad g_\nu = \frac{1\cdot 3 \cdots (2\nu - 1)}{2\cdot 4 \cdots 2\nu}, \qquad \nu = 1, 2, 3, \cdots. \tag{6.2.2}$$

Then $(\nu+1/2)g_\nu\lambda_\nu^{-1}$ is decreasing as $\nu$ increases since

$$\frac{\nu + 1/2}{\nu - 1/2}\,\frac{g_\nu}{g_{\nu-1}}\left(\frac{\lambda_\nu}{\lambda_{\nu-1}}\right)^{-1} = \frac{(\nu + 1/2)(\nu - 1)}{(\nu + 1)\nu} < 1. \tag{6.2.3}$$

Hence

$$(\nu + 1/2)g_\nu\lambda_\nu^{-1} > (M + 1/2)g_M\lambda_M^{-1}, \qquad \nu = 0, 1, \cdots, M - 1, \tag{6.2.4}$$

so that from (6.2.1)

$$g_M h_M \leqq \sum_{\nu=0}^{M-1} g_\nu(\lambda_\nu/\lambda_M)^{k/2-1}e^{(M-\nu)\gamma}V_\nu. \tag{6.2.5}$$

But

$$\lambda_\nu^{1/2} - \nu = (\nu(\nu + 1))^{1/2} - \nu = (1 + (1 + \nu^{-1})^{1/2})^{-1} \tag{6.2.6}$$

is positive and increasing with $\nu$ so that

$$\lambda_\nu/\lambda_M \leqq \exp\left\{2(\lambda_\nu^{1/2} - \lambda_M^{1/2})\lambda_M^{-1/2}\right\} \leqq \exp\left\{2(\nu - M)\lambda_M^{-1/2}\right\}. \tag{6.2.7}$$

Hence

$$g_M h_M \leqq \sum_{\nu=0}^{M-1} g_\nu V_\nu S^{\nu-M}, \qquad S = e^{(k-2)\lambda_M^{-1/2}-\gamma}. \tag{6.2.8}$$

From (6.1.1) we conclude that $S = S(M) \cong e^{H-\epsilon-\gamma} M$ as $M \to \infty$.

Writing

$$\zeta = (S + S^{-1})/2 \to \infty, \qquad M \to \infty, \tag{6.2.9}$$

we obtain by using a classical representation of Legendre polynomials[11]

$$P_\nu(\zeta) \geqq g_\nu S^\nu, \qquad \nu = 0, 1, 2, \cdots, \tag{6.2.10}$$

so that from (6.2.8) and (6.1.2)

$$g_M h_M \leqq S^{-M} \sum_{\nu=0}^{M-1} V_\nu P_\nu(\zeta) = S^{-M}(1 + \zeta)^N(1 + \zeta^m) - S^{-M}h_M P_M(\zeta). \tag{6.2.11}$$

The representation mentioned furnishes also $g_M h_M = 2^{-M}$ so that using (6.2.10) we again conclude

[11] See, for instance, G. Szegö, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23, 1939, p. 92, equation (4.9.4).

$$2^{1-M} \leqq S^{-M}(1+\zeta)^N(1+\zeta^m)$$

or

$$2^{1-M}(S/\zeta)^M \leqq (1+\zeta^{-1})^N(1+\zeta^{-m}) \leqq e^{\zeta^{-1}N}(1+\zeta^{-1}). \tag{6.2.12}$$

Now let $M\to\infty$. Then (cf. section 3.4)

$$2^{-M}(S/\zeta)^M = (1+S^{-2})^{-M} \to 1,$$
$$\zeta^{-1}M \to 2e^{-H+\epsilon+\gamma}$$

so that (3.4.8) follows. Consequently (3.1.1), (3.1.2) hold.

If the coefficients $c_\nu$ of the Legendre expansion (5.1.1) satisfy (3.1.1), then $f(x)$ must be analytic in an ellipse with foci at $-1$ and $+1$ the sum of the semi-axes of which is $e^T$ ([12]).

This establishes the proof of Theorem D.

## 7. Counterexamples

7.1. In this section we show that the conditions $N_k<k/\log k$ and $N_k<(k/\rho)^{1/\rho}/2$ of Theorems A and B cannot be replaced by $N_k=O(k)$ and $N_k=O(k^\alpha)$, respectively, where $\alpha>1/\rho$. I owe the necessary counterexamples to a suggestion of Professor Pólya.

7.2. The first assertion can be proved by considering the non-integral periodic function

$$f(x) = (1-2h\cos x+h^2)^{-1}, \qquad 0<h<1. \tag{7.2.1}$$

We see by mathematical induction that

$$f^{(k)}(x) = t_k(x)(1-2h\cos x+h^2)^{-k-1} \tag{7.2.2}$$

where $t_k(x)$ is a trigonometric polynomial of degree $k$. Consequently in this case $N_k\leqq 2k$.

7.3. Let $p>1$. The integral function

$$f(x) = \sum_{n=1}^{\infty} e^{-n^p}\cos nx \tag{7.3.1}$$

is([13]) of order $p/(p-1)$ and as we shall prove $N_k=O(k^{1/p})$. This furnishes, indeed, the desired counterexample by assuming $\alpha<1$ and choosing $p$ according to the conditions $1/\rho<1/p<\alpha$; then $\rho/(\rho-1)<p/(p-1)$.

Let $k$ be even. We apply Jensen's theorem to the function

$$f^{(k)}(x) = (-1)^{k/2}\sum_{n=1}^{\infty} n^k e^{-n^p}\cos nx \tag{7.3.2}$$

in the circle $|x|\leqq 2\pi$. Since

([12]) See Szegö, loc. cit., p. 238, Theorem 9.1.1.
([13]) See Footnote 8.

$$(7.3.3) \qquad |f^{(k)}(x)| \leqq \sum_{n=1}^{\infty} n^k e^{-n^p+2\pi n}$$

and

$$(7.3.4) \qquad |f^{(k)}(0)| = \sum_{n=1}^{\infty} n^k e^{-n^p}$$

we find for the number $N(\pi)$ of the roots of $f^{(k)}(x)$ in the circle $|x| \leqq \pi$

$$(7.3.5) \qquad 2^{N(\pi)} \sum_{n=1}^{\infty} n^k e^{-n^p} \leqq \sum_{n=1}^{\infty} n^k e^{-n^p+2\pi n}.$$

Obviously $N_k \leqq N(\pi)$.

In order to find a suitable bound for $N(\pi)$ let us consider the function $\lambda(\nu) = \nu^k e^{-\nu^p}$ of the continuous variable $\nu$, $\nu \geqq 1$. It is increasing for $\nu < \nu_0$ and decreasing for $\nu > \nu_0$ where

$$(7.3.6) \qquad \nu_0 = \nu_0(k) = (k/p)^{1/p}.$$

The maximum of $\lambda(\nu)$ is $\exp\{(k/p)\log(k/p) - k/p\}$.

The function $\lambda^*(\nu) = \nu^k e^{-\nu^p/2}$ assumes its maximum for $\nu_0^1 = \nu_0^1(k) = (2k/p)^{1/p}$.

Now let $\omega$ be fixed, $\omega > (2/p)^{1/p}$, $\log \omega - \omega^p/2 < -(\log p + 1)/p$. Then for $k \to \infty$

$$(7.3.7) \qquad \mathrm{I} = \sum_{n \leqq \omega k^{1/p}} n^k e^{-n^p+2\pi n} \leqq e^{2\pi\omega k^{1/p}} |f^{(k)}(0)|.$$

Further $\lambda^*(\nu)$ is decreasing for $\nu > \omega k^{1/p} > \nu_0'(k)$ so that

$$(7.3.8) \qquad \begin{aligned} \mathrm{II} &= \sum_{n > \omega k^{1/p}} n^k e^{-n^p+2\pi n} \leqq \lambda^*(\omega k^{1/p}) \sum_{n=1}^{\infty} e^{-(n^p/2)+2\pi n} \\ &= O(1) \exp\{k \log \omega + (k/p) \log k - \omega^p k/2\}. \end{aligned}$$

By use of the mean value theorem we find

$$(7.3.9) \qquad \log \lambda(\nu_0) \leqq \log \lambda([\nu_0] + 1) + Ck^{1-1/p}$$

where $C > 0$ is independent of $k$. But for large $k$

$$(7.3.10) \qquad \begin{aligned} k \log \omega + (k/p) \log k - \omega^p k/2 &< (k/p) \log (k/p) - k/p - Ck^{1-1/p} \\ &< \log \lambda([\nu_0] + 1), \end{aligned}$$

hence

$$(7.3.11) \qquad \begin{aligned} \mathrm{II} &= O(1)|f^{(k)}(0)|, \\ \mathrm{I} + \mathrm{II} &= O(1) e^{2\pi\omega k^{1/p}} |f^{(k)}(0)|. \end{aligned}$$

Consequently

$$2^{N(\pi)} = O(1)e^{2\pi\omega k^{1/p}} \tag{7.3.12}$$

from which $N(\pi) = O(k^{1/p})$ follows.

7.4. In case of an odd $k$ we have $f^{(k)}(0) = 0$. Then in Jensen's theorem $f^{(k)}(0)$ has to be replaced by

$$f^{(k+1)}(0) = (-1)^{(k+1)/2} \sum_{n=1}^{\infty} n^{k+1} e^{-n^p}. \tag{7.4.1}$$

The previous argument holds good except that $k$ has to be replaced by $k+1$.

Stanford University,
Stanford University, Calif.

# ON THE OSCILLATION OF DIFFERENTIAL TRANSFORMS. II

## CHARACTERISTIC SERIES OF BOUNDARY VALUE PROBLEMS

BY

EINAR HILLE

### INTRODUCTION

1. **Formulation of problem.** G. Pólya and N. Wiener [2](1) have recently made important contributions to the S. Bernstein problem concerning *the relation between the frequency of oscillation of derivatives of high order and the analytic character of the function.* Assuming $f(x)$ of period $2\pi$ and denoting the number of sign changes of $f^{(k)}(x)$ in the period by $N_k$, they show that restrictions in the rate of growth of $N_k$ when $k\rightarrow\infty$, imply that the high frequency terms in the Fourier series of $f(x)$ have "small" amplitudes. In particular, *if $N_k$ is bounded, $N_k\leqq N$ for all $k$,* then the high frequency terms are entirely missing and $f(x)$ *reduces to a trigonometric polynomial of degree at most $N/2$. Conversely, if $f(x)$ is a trigonometric polynomial of degree $K$, then $N_k=2K$ for all large $k$.* Their results are less precise when $N_k$ is unbounded. While it is likely that $N_k=O(k)$ is necessary and sufficient for analyticity of $f(x)$, this has not yet been proved, and the best they could do was to show that $N_k=o(k^{1/2})$ implies that $f(x)$ is an entire function.

For these and similar questions G. Szegö has devised a new method of attack, presented in the first paper of this series [4]. This method showed itself capable of giving more precise information when $N_k$ is unbounded. In particular, Szegö could show that $N_k<k(\log k)^{-1}$ implies that $f(x)$ is entire.

The present paper is also closely related to the paper of Pólya and Wiener, but proceeds in a different direction. We aim to preserve the essence of the methods developed by these writers and to apply them to a wider range of problems. There are several features in the investigation of Pólya and Wiener which suggest possible generalizations, in particular, the class of functions considered and the operations applied to them.

Let $T$ be an operator which takes functions $f(x)$ of a certain class $F$ into functions of the same class. Any function $u(x)$ of $F$ such that $Tu(x)=\lambda u(x)$ will be called a *characteristic function of $T$* corresponding to the *characteristic value* $\lambda$ and any formal series $\sum f_n u_n(x)$ will be called a *characteristic series of $T$* if its terms are characteristic functions.

In this terminology we can describe the investigation of Pólya and Wiener

Presented to the Society, November 22, 1941 under the title *On the oscillation of differential transforms.* II; received by the editors January 2, 1942.

(1) Numbers in square brackets refer to the references at the end of the paper.

as follows([2]). They are concerned with the differential operator $D^2$ and the characteristic functions of this operator determined by the periodic boundary value problem

$$(1.1)\qquad (D^2+\mu)y = 0, \qquad y(x+2\pi) = y(x).$$

Any function $f(x) \in C^{(\infty)}(-\infty, \infty)$, satisfying the same condition of periodicity $f(x+2\pi)=f(x)$, can be represented by a characteristic series of the operator,

$$(1.2)\qquad f(x) = (a_0/2) + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx),$$

to which the operator $D^2$ can be applied termwise as often as we please. They observe that for $\lambda>0$, $D^2-\lambda$ is an *oscillation preserving transformation* in the sense that the transform $(D^2-\lambda)f(x)$ has at least as many sign changes in the period as $f(x)$ has. This observation is used as follows.

Let $m$ be a positive integer and multiply the $n$th term of the series (1.2) by the $k$th power of the factor

$$(1.3)\qquad \left\{\frac{2mn}{m^2+n^2}\right\}^2.$$

A function $F(x, m, k; f)$ results which has at least as many sign changes in the period as $f^{(2k)}(x)$ since

$$(D^2-m^2)^{2k}F(x, m, k; f) = (2m)^{2k}f^{(2k)}(x).$$

On the other hand, for large values of $k$ the number of sign changes of $F(x, m, k; f)$ can be shown to be at least $2m$ provided the $m$th term is present in the original expansion (1.2). This is the basis for all their conclusions.

It is obvious from this formulation in what direction we are looking for extensions. Instead of the operator $D^2$ we shall consider a rather general linear differential operator $L$. In the present paper we restrict ourselves to second order operators satisfying certain conditions, but first or higher order operators would also be admissible. We define a set of characteristic functions of $L$ by a suitable boundary value problem for $L$ in the basic interval $(a, b)$ and consider the corresponding class of characteristic series, $F$ say, with the restriction that $L$ shall apply termwise to the series as often as we please. It turns out that the operator $L-\lambda$, $\lambda>0$, is always oscillation preserving in $(a, b)$ with respect to a suitable class of functions which includes $F$. Even the "root consuming factor" (1.3) has an obvious analogue in terms of characteristic values and the general procedure of Pólya and Wiener can be followed.

([2]) Actually Pólya and Wiener work with the operator $D$ and the corresponding characteristic functions exp $(nix)$. The "root consuming factor" in (1.3) is the square of their factor. The emphasis and terminology have been changed in order to bring out the generalizations.

It should be observed, however, that the method is not constrained to the consideration of characteristic series the terms of which are defined by boundary value problems and consequently orthogonal in the basic interval. The case of almost periodic functions was mentioned by Pólya and Wiener and a non-orthogonal characteristic series figures also in §2.11 of the present paper(³).

2. **Arrangement of material.** Chapter I is devoted to a study of oscillation preserving transformations defined by linear second order differential operators. The basic definitions are found in §1.1 while 1.2 contains a number of lemmas of the classical Sturmian type which are needed for the discussion. In §1.3 the operators $L$ are classified according to their behavior at the end points of the basic interval and to each of the four types considered we introduce function classes $B_\nu^{(k)}$ the elements of which satisfy, together with their $L$-transforms of order less than $k$, the corresponding types of boundary conditions. That $L-\lambda$, $\lambda>0$, is oscillation preserving with respect to $B_\nu^{(1)}$ is proved in §§1.4 to 1.7. Various extensions to functions of $L$ are discussed in §1.8 and the corresponding boundary value problems are introduced in 1.9. We call attention to the singular and semi-singular types which appear to be new, though many of the most useful orthogonal systems considered in analysis appear as solutions of such boundary value problems.

Chapter II brings the proof of the analogue of the Pólya-Wiener theorem on finite characteristic series. Here we place the discussion on a rather elaborate postulational basis to make up for our lack of knowledge of the existence and properties of solutions of the singular and semi-singular boundary value problems. We consider systems $S$ consisting of an operator $L$, a set of characteristic functions $\{u_n(x)\}$ with corresponding characteristic values $\{\mu_n\}$, and a basic interval $(a, b)$. We call the system *admissible* if it satisfies conditions $A_1$ to $A_6$ of §2.1. These are conditions which are well known to hold in the case of classical boundary value problems but which, conceivably, may fail in the case of singular ones. We also consider the class $F$ of admissible characteristic series $\sum f_n u_n(x)$ such that $\sum \mu_n^m |f_n| < \infty$ for all $m$. The convergence theory of such series is discussed in 2.2. The system $S$ is called *conservative* if the set $\{u_n(x), \mu_n\}$ belongs to an appropriate boundary

(³) There are no general results available relating to oscillation problems for non-orthogonal characteristic series. Existing evidence, meager as it is, seems to indicate that the situation is similar to the orthogonal case. In other words, if the frequency of oscillation of $L^k f(x)$ is bounded or has a finite limit inferior, then the frequency of oscillation of the components of $f(x)$ is similarly limited, the main difference being that we may now still have infinitely many components. "Characteristic integrals" can also be studied from this point of view by a suitable modification of the method. A first investigation of this type will be given by J. D. Tamarkin in a later paper in this series. The author wishes to use this opportunity to express his gratitude to his collaborators on the S. Bernstein problem, Professors G. Pólya, A. C. Schaeffer, G. Szegö, and J. D. Tamarkin, with whom he has had many profitable discussions of various points of his work during his stay at Stanford University.

value problem for $L$ in $(a, b)$ and it is shown that the results of Chapter I apply to conservative systems. In particular, $L-\lambda$, $\lambda>0$, and any real polynomial in $L$ with real positive roots are oscillation preserving in $(a, b)$ with respect to the class $F$. This is proved in §2.3 where we also discuss the relation between $F$ and the classes $B_\nu^{(\infty)}$ introduced in 1.3.

The main theorem is proved in 2.4. *If $S$ is conservative and $f(x)\in F$, then the assumption that the inferior limit of the number of sign changes of $L^kf(x)$ in $(a, b)$ is finite and equals $N$, implies that $f(x)$ is a linear combination of a finite number of characteristic functions $u_n(x)$, none of which can have more than $N$* (in an exceptional case possibly $N+1$) *sign changes in* $(a, b)$. This is the analogue of Theorem I of Pólya and Wiener. In §§2.5 and 2.6 we verify that the *classical boundary value problems lead to conservative systems*. In §§2.7 to 2.11 we give similar verifications for the systems of Legendre, Jacobi, Hermite, Weber-Hermite, and Laguerre, which correspond to singular boundary problems, and that of Bessel which is semi-singular. We call attention, in particular, to the characterization of ordinary polynomials by means of sign change properties given in Theorems 12, 13, and 14. It is analogous to the results of Pólya and Wiener for trigonometric polynomials quoted above.

Extensions to the case in which $N_k$ is unbounded are indicated briefly in §3.1 of the Appendix. The author has extended Theorem III of Pólya and Wiener under fairly general assumptions on the system, but the rather lengthy and complicated analysis is omitted here and the results are stated merely for the singular systems of §§2.7 to 2.10. It turns out that $N_k=o(k^{1/2})$ is again sufficient in order that the corresponding characteristic series shall converge in the finite complex plane and hence represent an entire function. The method of Szegö gives a better result, when it applies, which is to the Legendre and Jacobi cases.

## Chapter I. Oscillation preserving transformations

1.1. **Preliminary notions and formulas.** All functions considered in Chapters I and II are real functions of a real variable, defined in a finite or infinite interval $\langle a, b\rangle$ and having certain properties of continuity in $\langle a, b\rangle$. Here $\langle a, b\rangle$ stands for one of the four alternatives $(a, b)$, $(a, b]$, $[a, b)$, and $[a, b]$. The symbols $C^{(k)}\langle a, b\rangle$, with $k=0$, positive integer or $\infty$, refer to the usual continuity classes. Finally we denote the class of all functions real and holomorphic in $\langle a, b\rangle$ by $A\langle a, b\rangle$.

Let $g(x)\in C^{(0)}(a, b)$. We say that $g(x)$ *has $N$ changes of sign in $(a, b)$, if $(a, b)$ breaks up into exactly $N+1$ subintervals in each of which $g(x)$ keeps a constant sign, the signs being opposite in adjacent intervals.* The subintervals are in general not uniquely determined. The statement that the sign of $g(x)$ in $(x_1, x_2)$ is, for instance, positive is taken in the wide sense, that is, $g(x)\geqq 0$ and actually is greater than 0 in some subinterval of $(x_1, x_2)$. If $g(x)$ is periodic of period $b-a$, this definition should be slightly modified. We map the in-

terval on the circumference of a circle, identifying the end points. Here $N$ intervals of alternating signs determine $N$ sign changes. It is clear that $N$ must be even in the periodic case. If there is no finite $N$ with these properties, we say that $g(x)$ has infinitely many sign changes in $(a, b)$. The number of sign changes of $g(x)$ in $(a, b)$, finite or infinite, is denoted by $V[g(x)]$. The theorem of Rolle implies

LEMMA 1. *If $g(x) \in C^{(1)}(x_1, x_2)$ and $g(x) \to 0$ when $x \to x_1$ and when $x \to x_2$ but $g(x) \not\equiv 0$ in $(x_1, x_2)$, then $g'(x)$ has at least one sign change in $(x_1, x_2)$.*

Let $L$ denote the differential operator defined by

$$L[y] \equiv p_0(x)y + p_1(x)Dy + p_2(x)D^2y, \qquad D = d/dx, \tag{1.1.1}$$

where to start with the coefficients will be subjected to the following two assumptions which will be held fast throughout the paper:

$A_1$. $p_m(x) \in A(a, b)$, $m = 0, 1, 2$.

$A_2$. $p_0(x) \leqq 0$, $p_2(x) > 0$ *for* $a < x < b$.

For much of our work in §§1.2 to 1.7 it would be sufficient to assume merely $p_m(x) \in C^{(0)}(a, b)$, but any consideration involving repeated application of the operator requires additional restrictions of $p_m(x)$, so we may just as well assume analyticity from the start(4).

The self-adjoint form of $L$ is $L^*$ where

$$L^*[y] \equiv P(x)L[y] = D[P(x)p_2(x)Dy] + P(x)p_0(x)y, \tag{1.1.2}$$

$$P(x) = \frac{1}{p_2(x)} \exp\left\{\int^x \frac{p_1(t)}{p_2(t)}\,dt\right\}. \tag{1.1.3}$$

Here $P(x) > 0$ for $a < x < b$. If $p_1(x) \equiv 0$, we take $P(x) = 1/p_2(x)$.

Any solution of the differential equation

$$(L + \mu)y = 0, \tag{1.1.4}$$

$\mu$ constant, will be referred to as a *characteristic function of $L$ corresponding to the characteristic value $\mu$*. The reader should note that the terminology differs from that used in the Introduction according to which $-\mu$ rather than $\mu$ would be called the characteristic value. The present convention is preferable when one works with second order linear differential equations.

If $f(x) \in C^{(2)}(a, b)$, then $L[f]$ has a sense and $L[f] \in C^{(0)}(a, b)$. The differential transform $L[f]$ is the first *$L$-transform* of $f(x)$. The higher $L$-transforms are defined by recurrence:

$$L^k[f] = L[L^{k-1}[f]], \qquad L^0[f] = f. \tag{1.1.5}$$

If $f(x) \in C^{(2k)}(a, b)$, then $L^k[f]$ exists and belongs to $C^{(0)}(a, b)$. If convenient or desirable we drop the brackets or exhibit the variable. Thus $L^kf$, $L^kf(x)$,

(4) It should be observed in connection with $A_2$ that the theory goes through with only minor changes if $p_0(x)$ has merely a finite upper bound in $(a, b)$.

$L^k[f(x)]$, $L^k[f]$ all have the same meaning. The reader should observe that the symbol $L^kf(x_0)$, $a \leqq x_0 \leqq b$, denotes the value of $L^k[f]$ for $x=x_0$ and not the result of operating by $L^k$ on the constant $f(x_0)$.

DEFINITION. *Let $F$ be a subclass of $C^{(2)}(a, b)$ and let $\lambda$ be fixed real. Then $L-\lambda$ is said to be an oscillation preserving transformation in $(a, b)$ with respect to $F$ if*

$$V[(L - \lambda)f(x)] \geqq V[f(x)] \tag{1.1.6}$$

*for every $f(x) \in F$.*

It should be observed that there are always functions $f(x) \not\equiv 0$, satisfying (1.1.6). Thus if $\mu$ is fixed real and $y(x, \mu)$ is any solution of (1.1.1), then $(L-\lambda)\, y(x, \mu) = -(\lambda+\mu)\, y(x, \mu)$, so that (1.1.6) is trivially satisfied for every $\lambda \neq -\mu$. If $V[f(x)] = \infty$, (1.1.6) is understood to mean merely that the left-hand side is also infinite.

The basic formula in the discussion of the operator $L-\lambda$ is the factorization given by

$$(L - \lambda)f = \frac{p_2 W_2}{W_1} D\left\{\frac{W_1^2}{W_2} D\left[\frac{f}{W_1}\right]\right\}, \tag{1.1.7}$$

for which see L. Schlesinger [3, vol. I, p. 52]. Here, $W_1$ is a solution of the associated differential equation $(L-\lambda)\, y=0$, and $W_2$ is the Wronskian of $W_1$ and a second linearly independent solution of the same equation. It is permitted to assume that $W_2$ is real positive in $(a, b)$. The crucial point in the use of formula (1.1.7) lies in the choice of $W_1$ which we refer to as the *auxiliary solution.*

1.2. **General properties of the auxiliary solution.** We proceed to a discussion of the solutions of the associated differential equation[8]

$$(L - \lambda)y = 0, \qquad \lambda > 0, \tag{1.2.1}$$

in the interval $(a, b)$. Introducing

$$K(x) = P(x)p_2(x), \qquad G(x, \lambda) = P(x)[\lambda - p_0(x)], \tag{1.2.2}$$

we can rewrite the equation in the form

$$D[K(x)Dy] - G(x, \lambda)y = 0. \tag{1.2.3}$$

Under the assumption $A_2$, $K(x)$ and $G(x, \lambda)$ for $\lambda > 0$ are positive in $(a, b)$.

Two integrated forms of the equation will be useful in the following. First we have obviously

---

(8) The discussion follows the classical Sturmian pattern, but at least some of the required results do not appear to be available in a convenient form in the literature. The proofs are held down to a minimum.

$$(1.2.4)\qquad [K(x)y'(x)]_{x_1}^{x_2} = \int_{x_1}^{x_2} G(t, \lambda)y(t)dt.$$

Secondly, multiplying (1.2.3) by $y$ and integrating we get

$$(1.2.5)\qquad [K(x)y(x)y'(x)]_{x_1}^{x_2} = \int_{x_1}^{x_2} K(t)[y'(t)]^2dt + \int_{x_1}^{x_2} G(t, \lambda)[y(t)]^2dt.$$

We conclude from (1.2.5) that if $y(x)\not\equiv 0$ is a solution of (1.2.1) in $(a, b)$, then the product $y(x)\ y'(x)$ can vanish at most once in the interval. Hence the real solutions of (1.2.1) are of the following four types in $(a, b)$: (1) monotone solutions of constant sign, (2) solutions of constant sign having a maximum, (3) solutions of constant sign having a minimum, and (4) monotone solutions having a zero. These types are mutually exclusive and exhaust the possibilities. The fourth type is of no interest to us in the following and will be omitted from consideration.

The existence of unbounded solutions is vital in most of our discussion. We introduce the following notation:

$$(1.2.6)\qquad \begin{gathered} U(x, x_0) = \int_{x_0}^{x} \frac{dt}{K(t)}, \quad Q(x, x_0; \lambda) = \int_{x_0}^{x} G(v, \lambda)dv, \\ R(x, x_0; \lambda) = \int_{x_0}^{x} \frac{Q(t, x_0; \lambda)}{K(t)}dt. \end{gathered}$$

LEMMA 2. *Let $y(x)$ be the solution of* (1.2.1) *determined by the initial conditions $y(x_0)=1$, $y'(x_0)=s\geqq 0$, $a<x_0<b$. A necessary and sufficient condition that $y(x)\to\infty$ when $x\to b$ is that $R(x, x_0; \lambda)\to\infty$ when $x\to b$. If the latter condition is satisfied for a particular choice of $x_0$ and $\lambda$, then it holds for every $x_0$, $a<x_0<b$, and every $\lambda>0$. Moreover, if the condition holds, every solution of* (1.2.1) *such that $y(x)y'(x)$ is ultimately positive for approach to $b$ becomes infinite when $x\to b$. Similarly, if $R(x, x_0; \lambda)\to\infty$ when $x\to a$, then every solution with $y(x)y'(x)$ ultimately negative for approach to $a$ becomes infinite when $x\to a$.*

It is clear from the structure of $R(x, x_0; \lambda)$ that the condition is independent of $x_0$ and $\lambda$. We shall prove the lemma for fixed $x_0$ and $\lambda$ and consider only the case $x\to b$. The same method applies at the other end point. The lemma is an immediate consequence of

LEMMA 3. *Under the assumptions of Lemma* 2[6]

$$(1.2.7)\qquad S(x, x_0; \lambda) < y(x) < \exp\{S(x, x_0; \lambda)\}$$

---

[6] The inequality (1.2.7) does not give very precise information regarding the rate of growth of $y(x)$, but in a certain sense it is the best of its kind. The ratio $y(x)/S(x, x_0; \lambda)$ is bounded in the case of the Legendre operator $L=(1-x^2)D^2-2xD$, $a=-1$, $b=1$, for approach to the singular end points, while $y(x)\exp[-S(x, x_0; \lambda)]$ is bounded away from zero in the case of the Hermite operator $L=D^2-2xD$, $a=-\infty$, $b=\infty$. See §§2.7 and 2.9 below.

*for* $x_0 \leqq x < b$, *where*

$$(1.2.8) \qquad S(x, x_0; \lambda) = R(x, x_0; \lambda) + K(x_0)y'(x_0)U(x, x_0).$$

Putting $x_1 = x_0$ and $x_2 = u$ in (1.2.4) and noting that $y(x)$ is increasing and greater than 1 in the interval $(x_0, b)$, we get

$$K(u)y'(u) > K(x_0)y'(x_0) + Q(u, x_0; \lambda).$$

Dividing by $K(u)$ and integrating from $x_0$ to $x$, we obtain the first half of (1.2.7). But we have obviously also

$$K(u)y'(u) < K(x_0)y'(x_0) + Q(u, x_0; \lambda)y(u).$$

Dividing by $K(u)y(u)$, dropping $y(u) > 1$ in the first denominator on the right, and then integrating from $x_0$ to $x$, we get the second half of the inequality.

This shows that $y(x)$ becomes infinite when $x \to b$ if and only if $S(x, x_0; \lambda)$ has the same property. But both terms on the right in (1.2.8) are positive and a simple calculation shows that $U(x, x_0) \to \infty$ when $x \to b$ implies $R(x, x_0; \lambda) \to \infty$, but not vice versa. This completes the proof of the lemmas.

These inequalities have a relation to the transformation theory of the differential equation which is of some interest for the following. If we introduce a new independent variable in (1.2.1) by putting $u = U(x, x_0)$ and define $y(x) = Y(u)$, then the transformed differential equation is simply

$$(1.2.9) \qquad \frac{d^2Y}{du^2} - \frac{d^2R}{du^2} Y = 0,$$

where under our assumptions $d^2R/du^2 > 0$ in the interval $(A, B)$ which is the image of $(a, b)$ under the transformation. If, for instance, $B = \infty$, then it is perfectly trivial that every solution of (1.2.9), which is not positive monotone decreasing in $(A, B)$, becomes infinite with $u$. This transformation will be useful in the proof of the next lemma which is a comparison theorem of the classical Sturmian type.

LEMMA 4. *Let* $y(x; x_0, s, \lambda)$ *be the solution of* $(L - \lambda)y = 0$, $y(x_0) = 1$, $y'(x_0) = s \geqq 0$. (1) *For fixed* $x$, $x_0$, *and* $s$, $x_0 < x$, $y(x; x_0, s, \lambda)$ *is an increasing function of* $\lambda$. (2) *For fixed* $\lambda$, *the ratio of* $y(x; x_1, 0, \lambda)$ *to* $y(x; x_2, 0, \lambda)$, $a < x_1 < x_2 < b$, *lies between finite positive bounds depending upon* $x_1$ *and* $x_2$ *but not upon* $x$, $a < x < b$. (3) *For fixed* $x_0$ *and* $\lambda$, *the ratio of* $y(x; x_0, s_1, \lambda)$ *to* $y(x; x_0, s_2, \lambda)$, $0 \leqq s_1 < s_2$, *lies between finite positive bounds depending upon* $s_2$ *but not upon* $x$, $x_0 \leqq x < b$.

The first statement follows directly from the formula

$$K(x)\left[y_\mu(x)y_\lambda'(x) - y_\lambda(x)y_\mu'(x)\right] = (\lambda - \mu)\int_{x_0}^{x} P(t)y_\lambda(t)y_\mu(t)dt$$

with obvious notation. The second assertion lies slightly deeper, but follows from the expression for the Wronskian of two solutions $y_1(x)$ and $y_2(x)$ of the equation. Taking $y_1(x)=y(x;x_1,0,\lambda)$, $y_2(x)=y(x;x_2,0,\lambda)$, we get

$$K(t)W_2(y_1(t),\ y_2(t)) = K(x_1)y_2'(x_1) \equiv -C < 0.$$

Dividing through by $[y_2(t)]^2K(t)$ and integrating from $x_1$ to $x$ where $x_2<x$, we obtain

$$\frac{y_1(x)}{y_2(x)} = \frac{1}{y_2(x_1)} + C\int_{x_1}^{x} \frac{dt}{K(t)[y_2(t)]^2},$$

so the statement is proved for such values of $x$ if we can show the convergence of the integral when $x\to b$. This is trivial if the integral obtained by suppressing the factor $[y_2(t)]^2$ in the denominator is convergent. Hence we can assume that the function $U(x, x_0)$ of formula (1.2.6) tends to infinity when $x\to b$. Putting $u=U(x,x_1)$ and transforming the differential equation upon the form (1.2.9) we get

$$\int_{x_1}^{x} \frac{dt}{K(t)[y_2(t)]^2} = \int_0^u \frac{dv}{[Y_2(v)]^2}$$

with obvious notation. But $Y_2(v)$ is positive, concave upwards for $v>0$, and tends to infinity with $v$. Hence we can find a linear function $\alpha v+\beta$ with $\alpha>0$ such that $Y_2(v)>\alpha v+\beta$ for $v>v_0$. This proves the convergence of the integral and gives a finite upper bound for the ratio in the interval $(x_2, b)$ where the lower bound is unity. In the interval $(a, x_1)$ we simply interchange $y_1(x)$ and $y_2(x)$ and apply the same method. The interval $(x_1, x_2)$ is trivial. This completes the proof of (2). The same method can be used in proving (3).

1.3. **Boundary conditions.** We shall make no attempt to determine the maximal class with respect to which the operator $L-\lambda$ is oscillation preserving in $(a, b)$. It is likely to be a complicated and none too interesting problem. We shall instead specialize $L$ in various ways and determine certain associated classes of functions by means of appropriate boundary conditions. We shall consider four alternatives which by no means exhaust the field but which at least cover a large number of cases of well established interest.

$T_1$. Sturm-Liouville type. $p_m(x)\in A[a, b]$, $m=0, 1, 2$; $p_2(a)\neq 0$, $p_2(b)\neq 0$.

$T_2$. Periodic type. *Assumptions as under* $T_1$, *but in addition*, $K(a)=K(b)$, *that is*, $\int_a^b\{p_1(t)/p_2(t)\}\,dt=0$.

$T_3$. Singular type. $R(x, x_0;\lambda)\to\infty$ *when* $x\to a$ *and when* $x\to b$ *for some* $x_0$, $a<x_0<b$, *and* $\lambda>0$.

$T_4$. Semi-singular type. $p_m(x)\in A(a, b]$, $m=0, 1, 2$; $p_2(b)\neq 0$; *and* $R(x, b;\lambda)\to\infty$ *when* $x\to a$, $\lambda>0$.

In $T_3$ and $T_4$ the functions $R(x, x_0;\lambda)$ and $R(x, b;\lambda)$ are defined by (1.2.6). Lemma 2 shows that the value of $x_0$ is immaterial and that the condition holds for all $\lambda>0$ if it holds for a single one. In $T_4$ the roles of $a$ and $b$ can of course be interchanged.

With each operator $L$ of type $T_\nu$ we associate a set of classes $B_\nu^{(k)}\{L;\langle a, b\rangle\}$ of functions $f(x)$ satisfying appropriate boundary conditions. Here $k$ is a positive integer or infinity.

$B_1$. $B_1^{(k)}\{L;[a, b]\}\subset C^{(2k)}[a, b]; L^nf(a)=0, L^nf(b)=0, n=0, 1, \cdots, k-1$.

$B_2$. $B_2^{(k)}\{L;[a, b]\}\subset C^{(2k)}[a, b]; L^mf(a)=L^mf(b), m=0, 1, \cdots, k-1, k$; $DL^nf(a)=DL^nf(b), n=0, 1, \cdots, k-1$.

$B_3$. $B_3^{(k)}\{L;(a, b)\}\subset C^{(2k)}(a, b)$; $[L^nf(x)]/y(x; x_0, \lambda)\to 0$ *when* $x\to a$ *and when* $x\to b$ *for* $n=0, 1, \cdots, k-1$, *for some* $x_0$, $a<x_0<b$, *and arbitrarily small positive* $\lambda$.

$B_4$. $B_4^{(k)}\{L;(a,b]\}\equiv B_4^{(k)}\{L;C_1, C_2;(a,b]\}\subset C^{(2k)}(a,b]; [L^nf(x)]/y(x;b,\lambda)\to 0$ *when* $x\to a$ *for arbitrarily small positive* $\lambda$; $C_1L^nf(b)+C_2DL^nf(b)=0$, $C_1$, $C_2$ *fixed greater than or equal to* 0, *both conditions holding for* $n=0, 1, \cdots, k-1$.

Here $L^nf(a)$ and $DL^nf(a)$ are the values of $L^nf(x)$ and $DL^nf(x)$ at $x=a$. In $B_3$, $y(x; x_0, \lambda)=y(x; x_0, 0, \lambda)$ in the notation of Lemma 4, similarly in $B_4$ where $x_0=b$. By virtue of Lemma 4 we should expect that the value of $x_0$ is immaterial and that small positive values of $\lambda$ are the decisive ones. It is perfectly obvious that we could consider other classes of functions in connection with these operators. In particular, more general boundary conditions could be allowed at one end point in $B_1$. If $a$ and $b$ are interchanged in $B_4$, the sign of $C_2$ should also be changed. We merely mention these possibilities. Our main object in Chapter I will be to study the four listed types in some detail and to prove Theorem 1 and its various extensions.

THEOREM 1. *If the operator* $L$ *is of type* $T_\nu$ *and* $\lambda>0$, *then for every* $f(x)\in B_\nu^{(1)}\{L;\langle a, b\rangle\}$, $\nu=1, 2, 3, 4$, *we have*(7) $V[(L-\lambda) f(x)]\geqq V[f(x)]$, *that is*, $L-\lambda$ *is oscillation preserving in* $(a, b)$ *with respect to the corresponding class* $B_\nu^{(1)}\{L;\langle a, b\rangle\}$.

1.4. **Discussion of the Sturm-Liouville case.** This case is readily recognized and the proof of Theorem 1 is quite simple. We choose $W_1=y(x, \lambda)$ in (1.1.7) as the solution of the initial value problem $(L-\lambda) y=0$, $y(b, \lambda)=1$, $y'(b, \lambda)=0$. Formula (1.2.5) shows that $y(x, \lambda)>1$ in $(a, b)$.

The theorem is trivial if $V[f(x)]=\infty$. Suppose then that $V[f(x)]=N<\infty$. We can then find $N+2$ points $x_j$ where $a=x_1<x_2<\cdots<x_{N+1}<x_{N+2}=b$, such that $f(x_j)=0$ and $f(x)$ is not identically zero in anyone of the intervals $(x_j, x_{j+1})$. Since $y(x, \lambda)\geqq 1$, Lemma 1 shows that $D[f(x)/y(x, \lambda)]$ has at least one sign change in each of the $N+1$ intervals $(x_j, x_{j+1})$. Multiplication by the positive bounded factor $[y(x, \lambda)]^2/W_2$ does not change this situation and the derivative of the result by Lemma 1 has at least $N$ sign changes in $(a, b)$. Hence $V[(L-\lambda) f(x)]\geqq N$ and the theorem is proved.

If the boundary conditions in $B_1$ for $k=1$ be modified so that $f(a)=0$ is replaced by the condition $C_1f(a)-C_2f'(a)=0$, $C_1\geqq 0$, $C_2>0$, while the condi-

(7) $V[g]$ is to be computed according to the definition for periodic functions when $\nu=2$ but according to the main definition in the other cases. See §1.1, second paragraph.

tion $f(b)=0$ is left intact, the proof can still be carried through, but the choice of $y(x, \lambda)$ has to be modified accordingly. To each $f(x)$ of the class we determine a corresponding $y(x, \lambda)$ by the condition that it should have the same logarithmic derivative as $f(x)$ at $x=a$. Taking $y(a, \lambda)=1$ as is permissible, we still have $y(x, \lambda)>1$ in $(a, b)$. Then $D[f(x)/y(x, \lambda)]$ will be zero at $x=a$ instead of in the interior of $(a, x_1)$. It consequently still has $N+1$ zeros in $(a, b)$ and does not vanish identically between any consecutive pair of zeros. Thus $(L-\lambda) f(x)$ has $N$ sign changes at least, and the theorem is proved under the more general assumptions. The restriction imposed on the sign of the logarithmic derivative of $f(x)$ at $x=a$ is dictated solely by our concern that the corresponding $y(x, \lambda)$ shall be positive in $[a, b]$. If this condition is known to be satisfied, the restriction can be dropped. It is clear that modifying the boundary conditions at both end points meets with additional difficulties and this problem will not be considered here. It should be observed, however, that the case $f'(a)=0, f'(b)=0$, can be handled without difficulty.

1.5. **Discussion of the periodic case.** The name periodic case is to some extent a misnomer, but it is a customary designation for the corresponding type of boundary conditions and the case has close relations to periodicity in the usual sense. Moreover, it includes as a special instance the case in which $K(x)$ and $G(x, \lambda)$ are periodic with period $(b-a)$.

If $f(x)\in B_2^{(1)}\{L; [a, b]\}$, then $f(x)\in C^{(2)}[a, b]$, $f(a)=f(b)$, $f'(a)=f'(b)$, and $Lf(a)=Lf(b)$. We can then find a function $f^*(x)\in C^{(1)}(-\infty, \infty)$ such that $f^*(x+b-a)=f^*(x)$ and $f^*(x)=f(x)$ in $[a, b]$. The second derivative of $f^*(x)$ is continuous everywhere with the possible exception of $x\equiv a \pmod{(b-a)}$ where, however, right- and left-hand derivatives exist. Similarly $Lf(x)$ can be extended periodically as a continuous function and the extension agrees with $Lf^*(x)$.

The definition of $V[g(x)]$ given in §1.1 varied slightly according as $g(x)$ was defined only in $[a, b]$ or could be extended periodically as a continuous function with period $(b-a)$ outside of this interval. In the latter case the definition was such that the number of sign changes in the period would be independent of the choice of the end points. Actually the two definitions are always in agreement except in the case in which $g(a)=g(b)=0$ and $g(x)$ has an odd number, say $2K-1$, sign changes in the interior of the interval. In this case one definition would give $V[g(x)]=2K-1$ and the other $2K$, the zero at $x=a$ being counted as an additional sign change in the definition for periodic functions.

We now agree that if $\nu=2$ the definition for periodic functions shall be used in interpeting the $V$-symbols in Theorem 1. In other words, the inequality to be proved is actually

$$V[(L-\lambda)f^*(x)] \geqq V[f^*(x)]. \tag{1.5.1}$$

In the subsequent proof $V[g]$ refers to the non-periodic and $V[g^*]$ to the

periodic definition. The reader should note that $V[g] \leqq V[g^*] \leqq V[g]+1$ and $V[g^*]$ is always an even number.

For the proof we have to distinguish several subcases. Suppose first that $f(a)=f(b)=0$ but $V[f(x)]=V[f^*(x)]=2K$. The proof given in §1.4 applies without any change and gives $V[(L-\lambda)f(x)] \geqq 2K$ which in turn implies (1.5.1).

Suppose next that $f(a)=f(b)=0$ and $V[f(x)]=2K-1$, $V[f^*(x)]=2K$. We choose the same auxiliary solution $y(x, \lambda)$ as in the preceding case. By Lemma 1, $D[f(x)/y(x, \lambda)]$ has at least $2K$ sign changes in $(a, b)$. It follows that $V[(L-\lambda)f(x)] \geqq 2K-1$. Hence $V[(L-\lambda)f^*(x)] \geqq 2K$ and (1.5.1) follows.

Suppose finally that $f(a)=f(b)>0$ and $V[f(x)]=V[f^*(x)]=2K$. If $f'(a) \geqq 0$ we determine $y(x, \lambda)$ by the initial conditions $y(a, \lambda)=1$, $y'(a, \lambda)=f'(a)/f(a)$. If $f'(a)=f'(b)<0$, we take instead $y(b, \lambda)=1$, $y'(b, \lambda)=f'(b)/f(b)$. In either case $y(x, \lambda) \geqq 1$ in $[a, b]$. Then

$$[y(x, \lambda)]^2 \frac{d}{dx}\left\{\frac{f(x)}{y(x, \lambda)}\right\} = y(x, \lambda)f'(x) - y'(x, \lambda)f(x)$$

has at least $2K-1$ sign changes in $(a, b)$ and, in addition, vanishes at $x=a$ or $x=b$ depending upon the sign of $f'(a)$. It follows that $V[(L-\lambda)f(x)] \geqq 2K-1$ and $V[(L-\lambda)f^*(x)] \geqq 2K$. This completes the proof of Theorem 1 in the periodic case.

The proof is modelled upon that given by Pólya and Wiener for the case $L=D^2$.

1.6. **Discussion of the singular case.** This case is characterized by the presence of singular points of the differential equation at $x=a$ and $x=b$, sufficiently severe to cause the critical function $R(x, x_0; \lambda)$ to become infinite for $\lambda>0$. The class $B_3^{(1)}\{L; (a, b)\}$ consists of all functions $f(x) \in C^{(2)}(a, b)$ such that $f(x)/y(x; x_0, \lambda) \to 0$ when $x \to a$ and when $x \to b$ for arbitrarily small positive values of $\lambda$. Here $y(x; x_0, \lambda)$ is determined by $(L-\lambda)y=0$, $y(x_0)=1$, $y'(x_0)=0$, $a<x_0<b$. Let us first assess the influence of $x_0$ and $\lambda$ upon the determination of the class $B_3^{(1)}\{L; (a, b)\}$.

Suppose that $x_0$ and $\lambda$ are fixed and suppose $f(x) \in C^{(2)}(a, b)$, $f(x)/y(x; x_0, \lambda) \to 0$, $x \to a, b$. Denote the class of all such functions for the moment by $F^{(1)}(\lambda, x_0)$. By Lemma 4 the ratio of $y(x; x_1, \lambda)$ to $y(x; x_2, \lambda)$ is bounded away from zero and infinity in $(a, b)$. It follows that $f(x) \in F^{(1)}(\lambda, x_1)$ implies $f(x) \in F^{(1)}(\lambda, x_2)$ and vice versa so that $F^{(1)}(\lambda, x_0)$ is independent of $x_0$ and can be written simply $F^{(1)}(\lambda)$. Lemma 4 also asserts that $y(x; x_0, \lambda)$ is an increasing function of $\lambda$ in $x_0 \leqq x<b$. But in our case $s=0$ so that the argument given in Lemma 4, part (1), applies also to the interval $(a, x_0)$. Hence $f(x) \in F^{(1)}(\lambda_1)$ implies $f(x) \in F^{(1)}(\lambda_2)$ for $\lambda_1<\lambda_2$. In other words $F^{(1)}(\lambda_1) \subset F^{(1)}(\lambda_2)$ when $\lambda_1<\lambda_2$. Thus the cross section of all classes $F^{(1)}(\lambda)$ with $\lambda>0$ exists and equals $\lim_{\lambda\to 0} F^{(1)}(\lambda) \equiv F^{(1)}(+0)$. We can define in the same manner classes $F^{(k)}(\lambda)$ consisting of all functions $f(x)$ of $C^{(2k)}(a, b)$ for which $L^n f(x)/y(x; x_0, \lambda) \to 0$,

$x \to a, b, n=0, 1, \cdots, k-1$, as well as their cross section for $\lambda > 0$, $F^{(k)}(+0) = \lim_{\lambda \to 0} F^{(k)}(\lambda)$.

Since $y(x; x_0, 0)$ is well defined, so is the class $F^{(k)}(0)$ and it is clear that $F^{(k)}(0) \subset F^{(k)}(+0)$. Ordinarily these sets do not coincide because the sets $F^{(k)}(\lambda)$ are as a rule not continuous in $\lambda$. Even if they are continuous for $\lambda > 0$, they may very well lose this property for $\lambda = 0$. Simple examples can be given for both possibilities.

If $p_0(x) \equiv 0$, $y(x; x_0, 0) \equiv 1$ and there is no auxiliary solution (of constant sign) which becomes infinite at both end points for $\lambda = 0$. In this case $F^{(k)}(0)$ reduces simply to the subset of $C^{(2k)}(a, b)$ the elements of which satisfy the boundary conditions $L^n f(x) \to 0$, $x \to a, b, n = 0, 1, \cdots, k-1$. It is obvious that this set is a subset of every class $F^{(k)}(\lambda)$, $\lambda > 0$. If $p_0(x) \not\equiv 0$ and $R(x, x_0; 0) \to \infty$, $x \to a, b$, then $F^{(k)}(0)$ certainly contains elements which do not vanish on the boundary together with their $L$-transforms of order at most $k-1$.

These results allow us to formulate

THEOREM 2. $B_3^{(k)}\{L; (a, b)\} \equiv F^{(k)}(+0) \supset F^{(k)}(0)$.

The proof of Theorem 1 in the singular case can be given in a few lines. We choose $W_1 = y(x; x_0, \lambda)$ and proceed as in the Sturm-Liouville case, the only difference being that the points $x_i$ now figure as zeros of the continuous function $f(x)/y(x; x_0, \lambda)$ rather than of $f(x)$ which of course supplies the sign changes in $(a, b)$. Lemma 1 applies as before and gives $V[(L-\lambda) f(x)] \geqq N$(8).

1.7. **Discussion of the semi-singular case.** The discussion follows the same general pattern as in the singular case. The class $B_4^{(1)}\{L; C_1, C_2; (a, b]\}$ is defined as that subclass of $C^{(2)}(a, b]$ the elements of which satisfy at the singular end point $x = a$ the condition $f(x)/y(x; b, \lambda) \to 0$ for arbitrarily small $\lambda > 0$, while at the regular end point $x = b$ we have $C_1 f(b) + C_2 f'(b) = 0$ where $C_1 \geqq 0$, $C_2 \geqq 0$, $C_1 + C_2 > 0$ are fixed. The auxiliary solution $y(x; b, \lambda)$ satisfies the initial conditions $y(b) = 1$, $y'(b) = 0$.

Let us denote by $G^{(1)}(C_1, C_2; \lambda)$ the class of functions in $C^{(2)}(a, b]$ which satisfy these boundary conditions for a fixed $\lambda \geqq 0$. Lemma 4 ensures that the ratio of $y(x; b, 0, \lambda) = y(x; b, \lambda)$ to $y(x; b, -s, \lambda)$ for a fixed positive $s$ is bounded away from zero and infinity in $(a, b]$(9). Hence we have also $f(x)/y(x; b, -s, \lambda) \to 0$, $x \to a$, for any fixed $s > 0$, if $f(x) \in G^{(1)}(C_1, C_2; \lambda)$. As in §1.6 we show that $G^{(1)}(C_1, C_2; \lambda_1) \subset G^{(1)}(C_1, C_2; \lambda_2)$ for $\lambda_1 < \lambda_2$. We find that $G^{(1)}(C_1, C_2; +0) = \lim_{\lambda \to 0} G^{(1)}(C_1, C_2; \lambda)$ is the cross section of all classes $G^{(1)}(C_1, C_2; \lambda)$ for $\lambda > 0$. In a similar manner we define classes $G^{(k)}(C_1, C_2; \lambda)$ and $G^{(k)}(C_1, C_2; +0) = \lim_{\lambda \to 0} G^{(k)}(C_1, C_2; \lambda)$. Here $G^{(k)}(C_1, C_2; \lambda)$ is simply that subclass of $C^{(2k)}(a, b]$ consisting of functions $f(x)$ such that

(8) The same argument applies in case either end point should be regular or the condition $R(x, x_0; \lambda) \to \infty$ should fail to hold, provided $f(x)$ be constrained to vanish at the end point in question. Various intermediary types of operators are covered by this remark.

(9) A change of variable, replacing $x$ by $-x$, reduces the discussion to the case considered in Lemma 4.

$f(x), Lf(x), \cdots, L^{k-1}f(x)$ all belong to $G^{(1)}(C_1, C_2; \lambda)$. We have obviously $G^{(k)}(C_1, C_2; 0) \subset G^{(k)}(C_1, C_2; +0)$. We can sum up the result in

THEOREM 3. $B_4^{(k)}\{L; C_1, C_2; (a, b]\} = G^{(k)}(C_1, C_2; +0)$.

The proof of Theorem 1 in the semi-singular case follows the same lines as in the preceding cases. Suppose that $f(x) \in B_4^{(1)}\{L; C_1, C_2; (a, b]\}$. If $C_2 = 0$, that is, if $f(b) = 0$, we choose $y(x; \lambda) = y(x; b, \lambda)$ and proceed as in the singular case. If $C_2 \neq 0$, we set $s = C_1/C_2 = -f'(b)/f(b)$ and take $y(x, \lambda) = y(x; b, -s, \lambda)$. Putting $g(x) = f(x)/y(x, \lambda)$ we see that $g(x) \to 0$ when $x \to a$ and $g'(b) = 0$ since numerator and denominator of the fraction have the same logarithmic derivatives at $x = b$. The proof then proceeds as in the Sturm-Liouville case with generalized boundary conditions.

1.8. **Extensions.** The case $\lambda = 0$ figured briefly in §1.6. It is of some interest to determine function classes for which the operator $L$ itself is oscillation preserving. We arrive at the following result for the proof of which the reader will find the necessary material in the preceding sections.

THEOREM 4. *If the operator $L$ is of type* $T_\nu$, *there exists a class $F_\nu$ with respect to which $L$ is oscillation preserving in $(a, b)$. If $\nu = 1$ or 2 we have $F_\nu \supset B_\nu^{(1)}\{L; [a, b]\}$, while $F_3 \supset F^{(1)}(0)$ and $F_4 \supset G^{(1)}(C_1, C_2; 0)$.*

In the remainder of the paper we shall have to apply a given operator $L$ more than once to the functions under consideration. Here is where the classes $B_\nu^{(k)}\{L; \langle a, b\rangle\}$ with $k > 1$ are required. We note that if $f(x) \in B_\nu^{(k)}\{L; \langle a, b\rangle\}$ and $\lambda > 0$ then $(L - \lambda) f(x) \in B_\nu^{(k-1)}\{L; \langle a, b\rangle\}$. Repeated application of Theorem 1 leads to the following result.

THEOREM 5. *Let $\Pi_k(u)$ be a polynomial in $u$ of degree $k$, having real coefficients and real positive zeros. If $L$ is of type* $T_\nu$, *then $\Pi_k(L)$ is an oscillation preserving transformation in $(a, b)$ with respect to the corresponding class $B_\nu^{(k)}\{L; \langle a, b\rangle\}$.*

In particular, we can always allow the class $B_\nu^{(\infty)}\{L; \langle a, b\rangle\}$. It is obvious that $B_\nu^{(k)} \supset B_\nu^{(k+1)} \supset B_\nu^{(\infty)}$ and it can be shown that $B_\nu^{(\infty)}$ is never vacuous[10].

By virtue of Theorem 4 we can also allow the root $u = 0$ with arbitrary multiplicity, in cases $T_1$ and $T_2$ without restriction of the class and in cases $T_3$ and $T_4$ at least for the corresponding classes $F^{(k)}(0)$ and $G^{(k)}(C_1, C_2; 0)$. We can also extend in a different direction. We can allow operators of the form $E(L)$ where $E(u)$ is a suitably restricted entire function, provided we

[10] For $\nu = 1$ and 2, this follows from Theorems 7, 10, and 11 below. For $\nu = 3$ and 4 the statement is also obvious whenever the corresponding boundary value problems $P_3$ and $P_4$ of §1.9 have solutions. In more general cases, the following type of argument leads to functions having the desired properties. Suppose $\nu = 3$, $a$ and $b$ finite and at most poles of the coefficients. Then we can take any function of the form $\exp[-A(x-a)^{-2} - B(x-b)^{-2}]$, $A > 0$, $B > 0$. The modifications necessary in case $a$ or $b$ or both are infinite are obvious. Heavier singularities can be handled by stepping up on the exponential scale. The same type of functions will do for $\nu = 4$.

also restrict $f(x)$ to be analytic. The result, being of no importance for the following, is stated without proof.

THEOREM 6. *Let $E(u)$ be an entire function of order 1/2 and minimal type*[11]*, having real coefficients and real positive zeros. Let $L$ be of type* $T_r$. *Let $A_r\{L; \langle a, b\rangle\}$ be obtained from $B_r^{(\infty)}\{L; \langle a, b\rangle\}$ by replacing the requirement $f(x)\in C^{(\infty)}\langle a, b\rangle$ by $f(x)\in A\langle a, b\rangle$. Then $E(L)$ is an oscillation preserving transformation in $(a, b)$ with respect to the class $A_r\{L; \langle a, b\rangle\}$.*

What was said above regarding the root $u=0$ applies also, mutatis mutandis, to the case of an entire function.

1.9. **The associated boundary value problems.** With each operator $L$ of type $T_r$ there is an associated boundary value problem. We refer to the question of determining characteristic functions and characteristic values of the problem

$$(1.9.1) \qquad (L+\mu)u = 0, \qquad u(x) \in B_r^{(1)}\{L; \langle a, b\rangle\}.$$

Thanks to the analyticity assumptions for the coefficients of $L$ any solution must also have the property $u(x)\in A_r\{L; \langle a, b\rangle\}$. For the sake of clarity, we write out in full the four problems.

$P_1$. $(L+\mu)u=0$, $u(a)=0$, $u(b)=0$.

$P_2$. $(L+\mu)u=0$, $u(a)=u(b)$, $u'(a)=u'(b)$.

$P_3$. $(L+\mu)u=0$, $u(x)/y(x; x_0, \lambda)\to 0$, $x\to a, b$, for every $\lambda>0$.

$P_4$. $(L+\mu)u=0$, $u(x)/y(x; b, \lambda)\to 0$, $x\to a$, $C_1u(b)+C_2u'(b)=0$, $C_1\geqq 0$, $C_2\geqq 0$.

The problems $P_1$ and $P_2$ are classical boundary value problems of the Sturm-Liouville and periodic types, respectively. It is well known that these problems have solutions and the reader will find a short summary of the available information concerning the properties of the solutions, to the extent that is needed for our purposes, in §§2.5 and 2.6 below.

Boundary value problems of types $P_3$ and $P_4$ do not seem to have been discussed in the literature though a number of the best known special orthogonal systems used in analysis can be obtained as solutions of such problems. This is not the right place to develop a general theory of problems $P_3$ and $P_4$. We restrict ourselves here to pointing out the existence of the problems and will call attention to the special instances as they are encountered in Chapter II.

In the case of problems $P_1$ and $P_2$ there is in existence a well developed expansion theory. Thus, for instance, every function $f(x)\in B_1^{(1)}\{L; [a, b]\}$ can be represented by a uniformly convergent series in terms of characteristic functions of $P_1$. The same is true in the case of $P_2$. It is natural to expect that

---

(11) The statement means that $E(u)\exp(-\epsilon|u|^{1/2})\to 0$ when $|u|\to\infty$ for every $\epsilon>0$. It would be more precise to say that the order is at most 1/2 and if it equals 1/2, then the function is of minimal type.

a similar situation holds under fairly general circumstances also in the case of $P_3$ and $P_4$. A number of special instances are well known.

## CHAPTER II. FINITE CHARACTERISTIC SERIES

2.1. **Admissible systems.** In this chapter we shall start the study of the relationship between the infinitary behavior of the sequence $V[L^k f(x)]$ and the analytical nature of $f(x)$. This will be carried out under rather severe restrictions on $L$ and on $f(x)$. In part the restrictions are dictated by the nature of the problem, but they are also due to our lack of knowledge regarding the boundary problems $P_3$ and $P_4$ defined in §1.9. This makes it necessary for us to postulate the existence of a solution of the boundary problems involved with fairly regular properties of characteristic values and functions.

We consider first a system $S=S\{L, u_n(x), \mu_n; (a, b)\}$ consisting of an operator $L$, a set of characteristic functions $\{u_n(x)\}$ and corresponding characteristic values $\{\mu_n\}$, the interval being $(a, b)$. We say that $S$ *is admissible if it satisfies the assumptions* $A_1$ *to* $A_6$ *below and* $L$ *is of one of the types* $T_\nu$ defined in §1.3.

$A_1$. $p_m(x)\in A(a, b)$, $m=0, 1, 2$.

$A_2$. $p_0(x)\leqq 0$, $p_2(x)>0$, *for* $a<x<b$.

$A_3$. *The functions* $\{[P(x)]^{1/2}u_n(x)\}$ *form a real orthonormal system, complete in* $L_2(a, b)$.

$A_4$. $0<\mu_n\leqq\mu_{n+1}$. *The series* $\sum_1^\infty \mu_n^{-\alpha}$ *is convergent for some* $\alpha>0$.

$A_5$. *There exist constants* $\beta$ *and* $\gamma$ *and a non-negative function* $U(x)\in C^{(0)}(a, b)$, *such that*

$$|u_n(x)| \leqq \mu_n^{\beta} U(x), \qquad |u_n'(x)| \leqq \mu_n^{\gamma} U(x).$$

$A_6$. *For every fixed interval* $(c, d)$, $a\leqq c<d\leqq b$, $Z_n(c, d)$, *the number of zeros of* $u_n(x)$ *in* $(c, d)$, *tends to infinity with* $n$. $Z_n(a, b)$ *is finite and a never decreasing function of* $n$(12).

A number of admissible systems occurring in classical analysis will be exhibited in §§2.5 to 2.11 below.

We also consider a set $F=F\{L, u_n(x), \mu_n; (a, b)\}$ of characteristic series

$$\sum_{n=1}^{\infty} f_n u_n(x). \tag{2.1.1}$$

*This set will be called admissible if* $S$ *is admissible and*

$C_1$. $f_n$ *is a real for all* $n$,

$C_2$. $\sum_{n=1}^\infty \mu_n^k|f_n| < \infty$, $k=0, 1, 2, \cdots$.

This condition is obviously equivalent to the convergence of

(12) We have $Z_n(a, b)=V[u_n(x)]$ except possibly in the periodic case when we may have $Z_n(a, b)+1=V[u_n(x)]$.

$$\sum_{n=1}^{\infty} (\mu_n^m f_n)^2$$

for every integral value of $m$. In other words, the series

$$\sum_{n=1}^{\infty} (-\mu_n)^m f_n u_n(x) \tag{2.1.2}$$

represents a function $f_m(x)$ such that

$$[P(x)]^{1/2} f_m(x) \in L_2(a, b)$$

for $m=0, 1, 2, \cdots$. It is clear that $f_m(x)$ is obtained from $f(x)=f_0(x)$ by termwise operation with $L$ in the series (2.1.1). *A characteristic series is admissible if its coefficients satisfy* $C_1$ *and* $C_2$ *and* $S$ *is admissible. Such a series defines an admissible function.*

2.2. **Convergence in** $F$. An admissible series converges not merely in weighted mean square but also in the local sense.

LEMMA 5. *If* $f(x) \in F$, *then the series*

$$\sum_{n=1}^{\infty} f_n u_n(x), \qquad \sum_{n=1}^{\infty} f_n u_n'(x), \qquad f_n = \int_a^b P(t) u_n(t) f(t) dt,$$

*converge absolutely and uniformly in every fixed interval* $(a_1, b_1)$, $a<a_1<b_1<b$, *their sums being* $f(x)$ *and* $f'(x)$, *respectively. If the function* $U(x)$ *of* $A_5$ *can be taken equal to a constant, the convergence is uniform in* $[a, b]$.

The convergence properties follow from assumptions $A_5$ and $C_2$. The first series being convergent in $(a_1, b_1)$ both uniformly and in weighted mean square, we conclude that the uniform limit is equivalent to $f(x)$ and can be taken as the definition of $f(x)$ for all $x$. The sum of the uniformly convergent derived series is then obviously $f'(x)$.

LEMMA 6. *If* $f(x) \in F$, *so does* $L[f(x)]$ *and*

$$L[f(x)] = -\sum_{n=1}^{\infty} \mu_n f_n u_n(x). \tag{2.2.1}$$

For the proof we observe that the second derived series of $f(x)$ also converges absolutely and uniformly in $(a_1, b_1)$ and hence has the sum $f''(x)$. This follows from the identity

$$p_2(x) \sum_{n=j}^{k} f_n u_n''(x) = -p_1(x) \sum_{r=j}^{k} f_n u_n'(x) - \sum_{n=j}^{k} \mu_n f_n u_n(x) - p_0(x) \sum_{n=j}^{k} f_n u_n(x).$$

Hence $L[f(x)]$ exists and

$$L[f(x)] = \sum_{n=1}^{\infty} f_n L[u_n(x)] = -\sum_{n=1}^{\infty} \mu_n f_n u_n(x) \equiv f_1(x)$$

is an element of $F$.

It follows that all functions $L^m[f(x)] = f_m(x)$ exist and belong to $F$ whenever $f(x)$ does. Thus we can apply the operation $L$ as often as we please termwise to an admissible characteristic series and the result will stay in $F$. Such a series can also be differentiated termwise arbitrarily often, but it is not a priori obvious that the result is always in $F$, though this appears to be true in simple cases.

The class $F$ could evidently be characterized by descriptive properties. Its elements are real in $(a, b)$ and $[P(x)]^{1/2}f(x) \in L_2(a, b)$. $F$ is invariant under the operation $L$. It is a linear vector space with real multipliers and contains the basis $\{u_n(x)\}$. However, for our purposes it is simpler and more natural to start from the characteristic series.

2.3. **Conservative systems.** We need a couple of additional assumptions linking the classes $A_\nu\{L; \langle a, b\rangle\}$ of Theorem 6 with the systems $S$ and $F$. They read as follows.

$D_\nu$. *If $L$ is of type* $T_\nu$, *then* $u_n(x) \in A_\nu\{L; \langle a, b\rangle\}$ *for all* $n$.

$E_3$. *If* $\nu = 3$ *there exists a finite positive* $C(x_0, \lambda)$ *such that* $U(x) \leqq C(x_0, \lambda) y(x; x_0, \lambda)$ *for* $a < x < b$, $\lambda > 0$.

$E_4$. *If* $\nu = 4$ *there exists a finite positive* $C(\lambda)$ *such that* $U(x) \leqq C(\lambda)\, y(x; b, \lambda)$ *for* $a < x < b$, $\lambda > 0$.

It is worth while stating explicitly what $D_\nu$ amounts to in the various cases. Since $u_n(x)$ is a characteristic function of $L$, the denumerable infinity of boundary conditions entering into the definition of $A_\nu\{L; \langle a, b\rangle\}$ reduces to a single pair. We get:

$D_1$. $u_n(a) = 0$, $u_n(b) = 0$.

$D_2$. $u_n(a) = u_n(b)$, $u_n'(a) = u_n'(b)$.

$D_3$. $u_n(x)/y(x; x_0, \lambda) \to 0$, $x \to a, b$, *for all* $\lambda > 0$.

$D_4$. $u_n(x)/y(x; b, \lambda) \to 0$, $x \to a$, *for all* $\lambda > 0$, *and* $C_1 u_n(b) + C_2 u_n'(b) = 0$.

In other words, $D_\nu$ asserts the existence of a solution of the corresponding boundary value problem $P_\nu$ of §1.9 and that this solution is given by $\{u_n(x), \mu_n\}$. If $\nu = 1$ or 2, the function $U(x)$ of $A_5$ can be taken equal to a constant. This explains the absence of any conditions $E_1$ and $E_2$. In the two remaining cases we need an inequality between $U(x)$ and the auxiliary solution which is supplied by $E_3$ and $E_4$.

Definition. *An admissible system $S$ is called conservative if it satisfies the conditions* $D_\nu$ *and* $E_\nu$ *corresponding to its type* $T_\nu$.

Thus a conservative system satisfies conditions $A_1$ to $A_6$, one of the conditions $T_\nu$, $\nu = 1, 2, 3$ or 4, and the corresponding conditions $D_\nu$ and $E_\nu$.

THEOREM 7. *If* $S=S\{L,\ u_n(x),\ \mu_n;\ (a,\ b)\}$ *of type* $T_\nu$ *is conservative and* $F=F\{L,\ u_n(x),\ \mu_n;\ (a,\ b)\}$ *is the corresponding admissible set of functions, then* $F\subset B_\nu^{(\infty)}\{L;\ \langle a,\ b\rangle\}$. *If* $\nu=1$ *or* 2, $F=B_\nu^{(\infty)}$.

Suppose first that $\nu=1$ and $f(x)\in F$. We can then find a constant $U_1$ such that $|u_n(x)|\leqq U_1$, $a\leqq x\leqq b$, for all $n$[13]. Formula (1.2.4) with $\lambda=-\mu_n$ shows that $|u_n'(x)|\leqq\mu_n U_2$ for a suitably chosen constant $U_2$. Thus we can take $U(x)=U=\max\ (U_1,\ U_2)$, $\beta=0$, $\gamma=1$ in $A_5$. By Lemma 5 the characteristic series of $f(x)$ converges uniformly in $[a,\ b]$. Since every partial sum of the series vanishes for $x=a$ and $x=b$ by $D_1$, we have $f(a)=f(b)=0$. Further the first derived series converges uniformly in $[a,\ b]$ so that $f'(x)$ is also continuous in $[a,\ b]$. By Lemma 6, $L^kf(x)\in F$ for all $k$. This means that $f(x)$ has derivatives of all orders continuous in $[a,\ b]$ and $L^kf(a)=L^kf(b)=0$ for all $k$. Hence $f(x)\in B_1^{(\infty)}\{L;\ [a,\ b]\}$.

Suppose, conversely, that $f(x)\in B_1^{(\infty)}\{L;\ [a,\ b]\}$. This means that $f(x)\in C^{(\infty)}[a,\ b]$ and $L^kf(a)=L^kf(b)=0$ for all $k$. Since $L^kf(x)\in C^{(\infty)}[a,\ b]$ and vanishes at the end points, we have

$$L^kf(x) = \sum_{n=1}^{\infty} f_{n,k}u_n(x),$$

uniformly convergent in $[a,\ b]$. But here we can use the classical identity of Lagrange (for the notation, see formulas (1.1.2) and (1.2.2)):

$$gL^*[h] - hL^*[g] = D\{K(gh' - hg')\}.$$

If $g$ and $h$ belong to $B_1^{(\infty)}\{L;\ [a,\ b]\}$, integration from $a$ to $b$ gives

$$\int_a^b g(t)L^*[h(t)]dt = \int_a^b h(t)L^*[g(t)]dt$$

or

$$\int_a^b g(t)P(t)L[h(t)]dt = \int_a^b h(t)P(t)L[g(t)]dt$$

and by iteration

$$\int_a^b g(t)P(t)L^k[h(t)]dt = \int_a^b h(t)P(t)L^k[g(t)]dt$$

for every integer $k\geqq 0$. Putting in particular $g(x)=u_n(x)$, $h(x)=f(x)$ we get $f_{n,k}=(-\mu_n)^kf_n$. Since $f_{n,k}$ is real and $\sum f_{n,k}^2$ converges for every $k$, we see that the coefficients $f_n$ satisfy conditions $C_1$ and $C_2$. Hence $f(x)\in F$ and the theorem is proved for $\nu=1$.

The same type of argument applies if $\nu=2$, where of course periodicity plays the same role as vanishing on the boundary did when $\nu=1$.

[13] This follows from property (3) of §2.5.

Suppose now that $\nu=3$ and that $f(x)\in F$. In order to prove that $f(x)\in B_3^{(\infty)}\{L;\,(a,b)\}$ it is enough to show that $f(x)/y(x;\,x_0,\lambda)\to 0$, $x\to a$, $b$, for every $\lambda>0$; the $L$-transforms of $f(x)$ will then automatically satisfy the same conditions. But using $A_5$ and $E_3$ we have

$$\left|\frac{f(x)}{y(x;\,x_0,\lambda)}-\sum_1^N f_n\frac{u_n(x)}{y(x;\,x_0,\lambda)}\right|=\left|\sum_{N+1}^{\infty}f_n\frac{u_n(x)}{y(x;\,x_0,\lambda)}\right|$$
$$\leq C(x_0,\lambda)\sum_{N+1}^{\infty}|f_n|\mu_n^{\beta}.$$

By $C_2$ we can choose $N$ so large that the last member is less than any preassigned $\epsilon$, and by $D_3$ the finite series in the first member tends to zero when $x\to a$. This completes the proof.

Suppose finally $\nu=4$ and $f(x)\in F$. Since $y(x;\,b,\lambda)$ is bounded in $[a+\delta,\,b]$, $\delta>0$, for fixed $\lambda$, assumption $E_4$ shows that $U(x)$ is bounded in $[a+\delta,\,b]$. Hence by Lemma 5 the series for $f(x)$ and $f'(x)$ are uniformly convergent in $[a+\delta,\,b]$. The partial sums satisfy the boundary condition $C_1S_n(b)+C_2S_n'(b)=0$ for all $n$. Hence we have also $C_1f(b)+C_2f'(b)=0$ and the same boundary condition is satisfied by $L^kf(x)$ for all $k$. The proof that $f(x)/y(x;\,b,\lambda)\to 0$ when $x\to a$ for every $\lambda>0$ goes through as when $\nu=3$.

We cannot assert that $F=B_\nu^{(\infty)}$ when $\nu=3$ or 4. The following example disproves such a conjecture. We take for $L$ the Hermite-Weber operator $D^2-x^2$, $a=-\infty$, $b=\infty$; $\{u_n(x)\}$ is the set obtained by orthogonalizing and normalizing the Hermite polynomials and $\mu_n=2n+1$. It is shown in §2.9 that this system is conservative and of type $T_3$. If $f(x)=1$ then $L^kf(x)$ is an even polynomial of degree $2k$. Referring to formula (2.9.3) which gives the asymptotic behavior of $y(x;\,0,\lambda)$ for large $x$, we see that $f(x)=1$ belongs to $B_3^{(\infty)}\{L;\,(-\infty,\infty)\}$, but it does not satisfy the boundary conditions (2.9.7) for $k=0$; so it cannot belong to $F$.

Similarly $S\{D^2-x^2,\,u_{2n}(x),\,4n+1;\,(0,\infty)\}$ is a conservative system of type $T_4$, the regular boundary condition being $u'(0)=0$. Again $f(x)=1$ belongs to $B_4^{(\infty)}\{D^2-x^2;\,0,1;\,[0,\infty)\}$ but not to the corresponding class $F$ for which the singular boundary condition is still given by (2.9.7).

Combining Theorems 5 and 7 we get

THEOREM 8. *Let $S=S\{L,\,u_n(x),\,\mu_n;\,(a,b)\}$ be a conservative system and let $F$ be the corresponding set of admissible functions. Let $\Pi(u)$ be a polynomial in $u$ with real coefficients and real positive zeros. Then $\Pi(L)$ is an oscillation preserving transformation in $(a,b)$ with respect to $F$.*

2.4. **The main theorem.** We shall now prove

THEOREM 9. *Let $S=S\{L,\,u_n(x),\,\mu_n;\,(a,b)\}$ be a conservative system and let $F=F\{L,\,u_n(x),\,\mu_n;\,(a,b)\}$ be the corresponding set of admissible functions. Let $f(x)\in F$ and suppose that*

$$\liminf_{k\to\infty} V[L^k f(x)] = N < \infty. \tag{2.4.1}$$

*Then there exists an integer $M = M(N)$ such that $f_n = 0$ for $n > M$, that is*

$$f(x) = \sum_{n=1}^{M} f_n u_n(x). \tag{2.4.2}$$

*If all characteristic values are simple,*

$$V[u_M(x)] \leqq N, \tag{2.4.3}$$

*otherwise it is at most $N+1$. Conversely, if $f(x)$ is given by* (2.4.2), *then*

$$V[L^k f(x)] \geqq V[u_M(x)] \tag{2.4.4}$$

*for all large $k$.*

For the proof we employ the device of Pólya and Wiener [2] in suitable modification. To the given function $f(x) \in F$ with Fourier coefficients $f_n$ we form the auxiliary function

$$\Phi(x, k, m; f) = \sum_{n=1}^{\infty} \left\{ \frac{4\mu_m \mu_n}{(\mu_m + \mu_n)^2} \right\}^k f_n u_n(x) \tag{2.4.5}$$

where $k \geqq 0$, $m \geqq 1$ are arbitrary integers. We have $\Phi(x, 0, m; f) \equiv f(x)$. The multipliers are positive numbers less than or equal to 1 and equal to 1 only when $\mu_n = \mu_m$. Since every characteristic value is at most double, this means for at most two values of $n$. It follows that the coefficients of $\Phi$ also satisfy conditions $C_1$ and $C_2$ so that $\Phi \in F$. We can consequently apply the operator $(L - \mu_m)$ termwise as often as we please to the series (2.4.5). We find in particular that

$$(L - \mu_m)^{2k} \Phi(x, k, m; f) = (-4\mu_m)^k L^k f(x). \tag{2.4.6}$$

But $\mu_m > 0$ and by Theorem 8 $(L - \mu_m)^{2k}$ is an oscillation preserving transformation with respect to the class $F$ in $(a, b)$. Hence

$$N_k = V[L^k f(x)] \geqq V[\Phi(x, k, m; f)] \tag{2.4.7}$$

for every $k$ and $m$.

So far $m$ was arbitrary. We suppose now that $f_m \neq 0$. In order to take care of the slightly more complicated case in which there are double characteristic values, let us suppose $\mu_{m-1} = \mu_m$ and that also $f_{m-1} \neq 0$. We then write $\Phi = S_1 + S_2 + S_3$. Here $S_1$ is the finite sum from $n = 1$ to $n = m-2$, $S_2 = f_{m-1}u_{m-1}(x) + f_m u_m(x)$, while $S_3$ is the remainder. The trivial modification necessary if $\mu_m$ is simple is obvious. We shall estimate $S_1$ and $S_3$. The idea of the proof is to show that for sufficiently large values of $k$, $|S_1 + S_3|$ is dominated by $|S_2|$ at the maxima of the latter, provided that we restrict

ourselves to a fixed interior interval $(a_1, b_1)$, and that consequently the oscillatory properties of $\Phi$ in this interval are essentially the same as those of $S_2$. The latter, however, are regulated by assumption $A_6$ which ensures that the number of sign changes of $S_2$ in $(a_1, b_1)$ tends to infinity with $m$. This will lead to a contradiction for suitable values of $m$.

We consider now an arbitrary interior interval $(a_1, b_1)$, $a<a_1<b_1<b$. Let $B=\max U(x)$ for $a_1 \leqq x \leqq b_1$. Let

$$\delta = \max \frac{4\mu_m\mu_n}{(\mu_m + \mu_n)^2}$$

for $n \neq m-1$ and $m$ ($n\neq m$, if $\mu_m$ is simple). We have $\delta<1$. By assumption $A_5$

$$|S_1 + S_3| \leqq \delta^k \sum{}' |f_n u_n(x)| < \delta^k \sum_1^\infty |f_n| \mu_n^\beta \cdot U(x) \leqq \delta^k B \sum_1^\infty |f_n| \mu_n^\beta \equiv \delta^k T$$

for $a_1 \leqq x \leqq b_1$. Here the prime after the summation sign indicates that $n\neq m-1$ and $m$.

Let us write $Z_m(a_1, b_1) = j_m$. Assumption $A_6$ asserts that $j_m \to \infty$ with $m$. Since $\mu_m$ is a double characteristic value, $S_2(x) = f_{m-1}u_{m-1}(x) + f_m u_m(x)$ is a real solution of the differential equation $(L+\mu_m)\, y=0$. Consequently it has at least $j_m-1$ and at most $j_m+1$ zeros in $(a_1, b_1)$ by the classical oscillation theorems. Let the actual number be $i_m$. We can suppose without essential restriction of the generality that the zeros of $S_2(x)$ are interior to the interval $(a_1, b_1)$ and that $S_2(a_1)>0$. Then sgn $S_2(b_1) = (-1)^{i_m}$. Let the zeros occur at the points $x_\alpha$, $a_1<x_1<x_2< \cdots <x_{i_m}<b_1$. Let $\xi_\alpha$ be the uniquely determined point between $x_\alpha$ and $x_{\alpha+1}$ where $S_2'(x)=0$. If $S_2'(x)=0$ at a point in $(a_1, x_1)$, we denote this point by $\xi_0$; otherwise we set $\xi_0 = a_1$. Similarly $\xi_{i_m}$ is either the point where $S_2'(x)=0$ in $(x_{i_m}, b_1)$ or $b_1$ itself. We note that the points $\xi_0$ and $\xi_{i_m}$ are uniquely determined. Now let

$$\sigma = \min |S_2(\xi_\alpha)|, \qquad \alpha = 0, 1, \cdots, i_m.$$

We then choose $k$ so large that

$$\delta^k T < \sigma,$$

which is obviously possible. But this means that for such values of $k$ and $m$

$$\operatorname{sgn} \Phi(\xi_\alpha, k, m; f) = (-1)^\alpha, \qquad \alpha = 0, 1, \cdots, i_m.$$

Hence $\Phi(x, k, m; f)$ has at least $i_m$ sign changes in $(a_1, b_1)$ and a fortiori in $(a, b)$. Since $(L-\mu_m)^{2k}$ is an oscillation preserving transformation, formulas (2.4.6) and (2.4.7) show that $L^k f(x)$ also has at least $i_m$ sign changes in $(a, b)$ for all sufficiently large $k$. Hence $i_m \leqq N_k$ for all large $k$. But (2.4.1) asserts that $N_k = N$ for infinitely many values of $k$. This implies $i_m \leqq N$.

This is a contradiction, however, for large $m$ since $i_m$ tends to infinity with $m$. Since $i_m \geqq j_m - 1$, this gives a contradiction for $j_m > N+1$. We are thus

led to the conclusion that the characteristic series of $f(x)$ cannot contain any term $u_m(x)$ having more than $N+1$ zeros in $(a_1, b_1)$. But here $(a_1, b_1)$ is a perfectly arbitrary interior interval. It follows that the series of $f(x)$ contains no term $u_m(x)$ with $Z_m(a, b) > N+1$. If there are no multiple characteristic values, we can replace $N+1$ by $N$ since then $j_m = i_m$. In the case of double characteristic values, however, it would seem possible for the finite sum to end with two terms corresponding to the same characteristic value, either term having $N+1$ zeros while their sum has only $N$ zeros. Whether or not this exceptional case can ever arise must be left an open question.

This argument proves formula (2.4.3) except in the periodic case. Here $N=2K$ is an even integer. Further $Z_n(a, b) \leqq V[u_n(x)] \leqq Z_n(a, b)+1$ and $V[u_n(x)]$ is even. If there are no double characteristic values then the inequality $Z_M(a, b) \leqq N = 2K$ implies $V[u_M(x)] \leqq 2K$ and formula (2.4.3) is proved. If however, $\mu_M = \mu_{M-1}$, then the previous proof shows that $S_2(x)$ cannot have more than $2K$ sign changes in $(a, b)$ and hence $V[S_2(x)] \leqq 2K$. Now $u_{M-1}(x)$, $u_M(x)$, and $S_2(x)$ are solutions of the same differential equation $(L+\mu)\,u=0$ for $\mu=\mu_M=\mu_{M-1}$. Hence the three quantities $V[u_{M-1}(x)]$, $V[u_M(x)]$ and $V[S_2(x)]$ can differ by at most one unit and, being even integers, they must consequently be equal. This shows that $V[u_M(x)] \leqq 2K$ and formula (2.4.3) is proved.

Suppose, conversely, that $f(x)$ is a finite sum of characteristic functions given by (2.4.2). We choose $m=M$ and form $\Phi(x, k, M; f)$. For $(a_1, b_1)$ we take an interval containing all zeros of $u_M(x)$ or $S_2(x)$ as the case may be. Proceeding as above, we see that $\Phi(x, k, M; f)$ has at least as many sign changes in $(a_1, b_1)$ as the last term or group of terms has for large values of $k$. Combining with (2.4.7) we see that (2.4.4) holds. The argument is evidently also valid in the periodic case. It is often possible to exclude the sign of inequality both in (2.4.3) and in (2.4.4). This completes the proof of Theorem 9.

Pólya and Wiener proved that if $f(x)$ is periodic and $V[D^{2k}f(x)]$ is bounded with respect to $k$, then the Fourier series of $f(x)$ cannot contain any high frequency terms. Theorem 9 shows that this result has analogues for general orthogonal series defined by boundary value problems relating to linear second order differential equations, the operator $D^2$ being replaced by $L$.

In §§2.5 to 2.11 below we shall give special instances of the theory.

2.5. **Sturm-Liouville operators.** We have the following simple results(14):

THEOREM 10. *Let* $p_m(x) \in A[a, b]$, $p_0(x) \leqq 0$, $p_2(x) > 0$, $a \leqq x \leqq b$. *Let* $\{u_n(x)\}$ *and* $\{\mu_m\}$, $n=0, 1, 2, \cdots$, *be the sets of characteristic functions and corresponding characteristic values of the boundary value problem*

$$(L+\mu)u = 0, \qquad u(a) = 0, \qquad u(b) = 0.$$

(14) We state the assumptions in Theorems 10 and 11 explicitly since they are so simple. Formulations in terms of the previous postulates are given below.

*Then* $S\{L, u_n(x), \mu_n; (a, b)\}$ *is a conservative system. If* $f(x) \in B_1^{(\infty)}\{L; [a, b]\}$, *that is, if* $f(x) \in C^{(\infty)}[a, b]$ *and* $L^n f(a) = 0$, $L^n f(b) = 0$ *for* $n = 0, 1, 2, \cdots$, *and*

$$\liminf_{k\to\infty} V[L^k f(x)] = N,$$

*then*

$$f(x) = \sum_{n=0}^{N} f_n u_n(x).$$

We are assuming the validity of $A_1$, $A_2$, $T_1$, and $D_1$ and have first to show that they imply $A_3$ to $A_6$. Now this is the classical Sturm-Liouville system except for the restrictive assumptions of analytical coefficients which are unnecessary in the boundary value problem but desirable for our special needs. Referring to the literature for proofs (see for instance E. L. Ince [1, §§10.61, 10.7, and 11.4]), we list the following properties of the solutions. We put

$$z = (1/\omega)\int_a^x [p_2(t)]^{-1/2}dt, \qquad \omega = (1/\pi)\int_a^b [p_2(t)]^{-1/2}dt,$$

$$\rho_n^2 = \omega^2 \mu_n.$$

Then:

(1) The characteristic values are real, positive, and simple.

(2) $\rho_n = n + 1 + O(1/n)$.

(3) $u_n(x) = A_n[P^2(x)\,p_2(x)]^{-1/4}\{\sin(\rho_n z) + O(1/n)\}$, where $A_n$ is a normalizing factor, independent of $x$ and bounded with respect to $n$[15].

(4) $V[u_n(x)] = n$.

(5) $\{[P(x)]^{1/2}u_n(x)\}$ is complete in $L_2(a, b)$.

These properties show that conditions $A_3$ to $A_6$ are amply satisfied. Thus $S$ is a conservative system and Theorem 9 holds for the corresponding set $F$ of admissible series. It was shown in Theorem 7, however, that $F = B_1^{(\infty)}\{L; [a, b]\}$. Finally, $M(N) = N$ by virtue of property (4). This completes the proof of Theorem 10.

The same result is valid for more general boundary conditions, for example, $u(a) = 0$, $C_1u(b) + C_2u'(b) = 0$, $C_1 \geqq 0$, $C_2 > 0$ and, for $u'(a) = 0$, $u'(b) = 0$.

2.6. **Periodic operators.** Here we also have simple results.

THEOREM 11. *Let* $p_m(x) \in A[a, b]$, $m = 0, 1, 2$; $p_2(a) \neq 0$, $p_2(b) \neq 0$, $K(a) = K(b)$. *Let* $\{u_n(x)\}$ *and* $\{\mu_n\}$ *be the sets of characteristic functions and corresponding characteristic values of the boundary value problem*

$$(L + \mu)u = 0, \qquad u(a) = u(b), \qquad u'(a) = u'(b).$$

*Then* $S\{L, u_n(x), \mu_n; (a, b)\}$ *is a conservative system. If* $f(x) \in B_2^{(\infty)}\{L; [a, b]\}$, *that is, if* $f(x) \in C^{(\infty)}[a, b]$ *and* $L^n f(a) = L^n f(b)$, $DL^n f(a) = DL^n f(b)$, $n = 0, 1,$

[15] $A_n \to \{2/(\pi\omega)\}^{1/2}$ when $n \to \infty$.

2, $\cdots$, *and if*

$$\liminf_{k\to\infty} V[L^k f(x)] = 2K,$$

*then*

$$f(x) = \sum_{n=0}^{2K} f_n u_n(x).$$

Here we assume $A_1$, $A_2$, $T_2$, and $D_2$, and want to conclude that $A_3$ to $A_6$ hold. In the present section $V[g]$ is to be determined according to the definition for periodic functions. The available information concerning the solutions of the boundary value problem is quite precise in the case of equations with periodic coefficients and only slightly less so in the general case. We refer to E. L. Ince [1, §§10.8, 10.81, and 11.4], where the reader will find further references to the literature. In the notation of the preceding section we obtain:

(1) The characteristic values are real, non-negative but need not be simple.

(2) $\rho_n = [(n+1)/2]+O(1)$.

(3) $|u_n(x)| \leqq U, a \leqq x \leqq b, n=0, 1, 2, \cdots$.

(4) $V[u_0(x)]=0$, $V[u_{2m-1}(x)]=V[u_{2m}(x)]=2m$.

(5) $\{[P(x)]^{1/2}u_n(x)\}$ is complete in $L_2(a, b)$.

If $K(x)$ and $G(x, \lambda)$ are even periodic functions of period $(b-a)$, the remainder term in (2) can be replaced by $O(1/n)$ and (3) can be replaced by formulas of type (3), in §2.5 with sine replaced by cosine when $n$ is even[16]. The properties as listed are, however, more than sufficient to prove that $A_3$ to $A_6$ are satisfied so that $S$ is a conservative system. Since $F=B_2^{(\infty)}\{L; [a, b]\}$ by Theorem 7, and $M(N)=2K$ by (4), Theorem 11 is proved.

The simplest of all operators satisfying the conditions of Theorem 11 is $L=D^2$. In this case the theorem reduces to Theorem I of Pólya and Wiener. A less trivial instance is given by the operator of Mathieu

$$L = D^2 - (A + B\cos 2x)$$

to which corresponds expansions in terms of the *functions of the elliptic cylinder*.

The remaining sections of the chapter will be devoted to special instances of singular and semi-singular operators.

2.7. **The Legendre operator.** We consider the case

$$L[y] = (1 - x^2)D^2y - 2xDy, \qquad a = -1, \qquad b = +1. \tag{2.7.1}$$

The end points of the interval are singular and we find that $R(x, 0; \lambda) = -(\lambda/2)\log(1-x^2)\to\infty$ when $x\to\pm 1$. It follows that the problem is of type $T_3$. The corresponding singular boundary value problem

(16) It is not difficult to show that similar formulas hold also in the general case considered in §2.6. We have merely to replace $\rho_n x$ by $\rho_n x+\phi_n$ where $\phi_n$ is a suitable phase angle, determinable with an error which is $O(1/n)$. Property (3) is an immediate consequence of such formulas.

$$(2.7.2) \qquad D[(1-x^2)Du]+\mu u=0, \quad u(x)/\log(1-x^2)\to 0, \quad x\to\pm 1$$

has as its characteristic functions the Legendre polynomials $P_n(x)$ with corresponding characteristic values $n(n+1)$, $n=0, 1, 2, \cdots$ ([17]). We take $u_n(x)=(n+(1/2))^{1/2}P_n(x)$. We shall prove that the system $S\{L, u_n(x), \mu_n; (-1, 1)\}$ is conservative. We know to start with that $A_1$, $A_2$, $T_3$, and $D_3$ are satisfied. It is well known that $A_3$ holds and so does obviously $A_4$, except for that fact that the least $\mu_n$ is zero. This is immaterial, however([18]). We have $|P_n(x)|\leqq 1$, $|P_n'(x)|\leqq n(n+1)/2$, so that $A_5$ and $E_3$ are satisfied. Further the $n$ zeros of $P_n(x)$ are all located in $(-1, 1)$ and the maximal distance between consecutive zeros is $O(1/n)$ so that $A_6$ is valid. Thus the system is actually conservative and Theorem 9 holds for the corresponding class $F$.

We have now to determine what functions are represented in $[-1, 1]$ by expansion of the form

$$\sum_{n=0}^{\infty} a_nP_n(x), \qquad \sum_{1}^{\infty} n^m|a_n| < \infty$$

for all $m$, $a_n$ being real. It is obvious that the series as well as all derived series converge uniformly in $[-1, 1]$ so that $f(x)\in C^{(\infty)}[-1, 1]$. Conversely, if $f(x)\in C^{(\infty)}[-1, 1]$ so do all its $L$-transforms. From this we conclude readily that $f(x)\in F$. Hence we have shown that

$$(2.7.3) \qquad F\{D(1-x^2)D, (n+(1/2))^{1/2}P_n(x), n(n+1); (-1, 1)\}\equiv C^{(\infty)}[-1, 1].$$

This fact gives the following formulation of Theorem 9 for the Legendre operator:

THEOREM 12. *If* $L=(1-x^2)D^2-2xD$, $f(x)\in C^{(\infty)}[-1, 1]$, *and* $\liminf_{k\to\infty} V[L^kf(x)]=N$, *then* $f(x)$ *is a polynomial of degree* $N$. *Conversely, every real polynomial of exact degree* $N$ *has the property* $V[L^kf(x)]=N$ *for all large* $k$.

In order to prove the converse, we merely express the given polynomial in terms of Legendre polynomials. The expression will involve the term $P_N(x)$ with a coefficient different from zero. By (2.4.4) $V[L^kf(x)]\geqq N$ for all large $k$. Since $L^kf(x)$ is also a polynomial of degree $N$, we must have $V[L^kf(x)]=N$ for all large $k$.

Theorem 12 has also been proved by Szegö [4, Theorem C] by a different method.

---

([17]) The proof of this statement goes as follows. The only solution $u(x)$ satisfying the boundary condition at $x=1$ is a multiple of $F(a+1, -a, 1, (1-x)/2)$ where $a(a+1)=\mu$. This solution becomes logarithmically infinite at $x=-1$ unless $a$ is an integer when it reduces to $P_n(x)$.

([18]) The fact that $\mu_0=0$ means merely that the case $N=0$ is not covered by Theorem 9. But if $N=0<1$ we can still conclude that $f(x)=a+bx$ and since $L^k(a+bx)=(-2)^kbx$, we must have $b=0$. Hence Theorem 12 is also valid for $N=0$. A similar argument takes care of the other cases encountered below in which the least characteristic value is zero.

2.8. **Jacobi operators.** Analogous results can be proved for the Jacobi operator

(2.8.1) $$L = (1 - x^2)D^2 + [\beta - \alpha - (\alpha + \beta + 2)x]D$$

for the interval $(-1, 1)$. Here the end points are again singular. A simple calculation shows that $R(x, 0; \lambda) \to \infty$ when $x \to 1$ if and only if $\alpha \geqq 0$, and when $x \to -1$ if and only if $\beta \geqq 0$. Thus the problem is of type $T_3$ if and only if $\alpha \geqq 0, \beta \geqq 0$. We shall suppose $\alpha > 0, \beta > 0$, since the limiting case $\alpha = 0, \beta = 0$, reduces to Legendre's operator(19). The corresponding singular boundary value problem can then be formulated as follows:

(2.8.2) $$(L + \mu)u = 0, \quad u(x)(1 - x)^\alpha(1 + x)^\beta \to 0, \quad x \to \pm 1,$$

since the calculation shows that $(1-x)^\alpha(1+x)^\beta\, y(x; 0, \lambda)$ is bounded away from zero and infinity in $(-1, 1)$. The solutions are given by the Jacobi polynomials $u_n(x) = A_n P_n^{(\alpha,\beta)}(x)$, $\mu_n = n(n+\alpha+\beta+1)$, where $A_n$ is a normalizing factor. The reader will find in the treatise by Szegö [5, §§3.1, 7.32, and 8.9], the necessary information regarding Jacobi polynomials required to show that $S\{L, u_n(x), \mu_n; (-1, 1)\}$ is a conservative system. We show as in §2.7 that the corresponding set $F$ of admissible functions is identical with $C^{(\infty)}[-1, 1]$.

It follows that Theorem 12 remains valid if we replace the Legendre operator by the general Jacobi operator (2.8.1) provided $\alpha \geqq 0, \beta \geqq 0$. Professor Szegö has kindly informed me that the theorem actually remains true for $\alpha > -1, \beta > -1$ and that this can be proved both by his method used in [4] and by a suitable modification of mine. We note, in particular, the case $\alpha = \beta = -1/2$ which leads to the polynomials of Tchebycheff. By his method Szegö is also able to prove that if $\alpha$ and $\beta$ are arbitrary real numbers, then the assumptions $f(x) \in C^{(\infty)}[-1, 1]$ and $\lim \inf_{k\to\infty} V[L^k f(x)] = N$ imply that $f(x)$ is a polynomial of degree at most $N + M(\alpha, \beta)$ where $M(\alpha, \beta)$ is an integer depending only upon $\alpha$ and $\beta$. Detailed proofs will be presented in a later note in this series.

2.9. **The Hermite and Hermite-Weber operators.** We consider next the two operators

(2.9.1) $$L_1 = D^2 - 2xD, \qquad L_2 = D^2 - x^2,$$

which we refer to as the Hermite and Hermite-Weber operators respectively. The interval $(a, b)$ will be $(-\infty, \infty)$. Since

(2.9.2) $$e^{x^2/2}L_2(e^{-x^2/2}y) = (L_1 - 1)y,$$

the two operators can be treated simultaneously. The Hermite-Weber case

(19) The cases $\alpha = 0, \beta > 0$ and $\alpha > 0, \beta = 0$ can also be handled by the same method. One of the powers occurring in (2.8.2) should then be replaced by a logarithm.

is easier to handle directly, but the final result is more striking if expressed in terms of the Hermite operator.

We concentrate the attention on $L_2$. The point at infinity is singular and $R(x, 0; \lambda)\to\infty$ when $|x|\to\infty$ for $\lambda>0$. The problem, therefore, is of type $T_3$. The function $y(x, 0; \lambda)$ is a constant multiple of $D_n(2^{1/2}x)+D_n(-2^{1/2}x)$ where $D_n$ is the *parabolic cylinder function* of Whittaker and $n=-(1+\lambda)/2$. For large values of $|x|$ we have consequently (cf. E. T. Whittaker and G. N. Watson [6, §16.52])

$$(2.9.3)\qquad y(x, 0; \lambda) = B(\lambda)\,|\,x\,|^{(\lambda-1)/2}\exp\,[\,x^2/2]\{1+o(1)\}.$$

The singular boundary value problem

$$(2.9.4)\quad (D^2+\mu-x^2)u = 0,\ u(x)\,|\,x\,|^{(1-\lambda)/2}\exp\,[-\,x^2/2]\to 0,\ |\,x\,|\to\infty,$$

has for solutions the Weber-Hermite functions[20]

$$u_n(x) = [\pi^{1/2}2^n n!]^{-1/2}(-1)^n e^{x^2/2}D^n(e^{-x^2}),\qquad \mu_n = 2n+1.$$

It is well known that this system is complete in $L_2(-\infty, \infty)$. Condition $A_5$ is fulfilled since[21]

$$(2.9.5)\qquad |\,u_n(x)\,| \leqq B_1,\qquad |\,u_n'(x)\,| \leqq B_2 n^{1/2}.$$

All zeros of $u_n(x)$, zeros of the $n$th Hermite polynomials, are real and located in the interval $(-\mu_n^{1/2}, \mu_n^{1/2})$. They are densest towards the center of the interval, but the minimum distance between consecutive zeros is of the same order as the average distance. It follows that the conditions $A_3$ to $A_6$ and $E_3$ are satisfied. Hence $S$ is a conservative system and Theorem 9 applies to the corresponding set $F$.

The determination of the class $F$ is much more laborious than in the Legendre case. We know that

$$f(x) = \sum_{n=0}^{\infty} f_n u_n(x),\qquad \sum_{n=1}^{\infty} n^m |\,f_n\,| < \infty$$

for all $m$. The series is uniformly convergent in the infinite interval by virtue of (2.9.5) and the terms tend to zero as $|x|\to\infty$. It follows that $f(x)\in C_0^{(0)}[-\infty, \infty]$ where the subscript 0 indicates that $f(x)\to 0$ when

(20) For the proof it is enough to observe that the only solution which satisfies the boundary condition when $x\to\infty$ is a multiple of $D_\nu(2^{1/2}x)$, $\nu=(\mu-1)/2$, and that this solution, as is seen from its asymptotic representation, does not satisfy the boundary condition for $x\to-\infty$ unless $\mu$ is a positive odd integer.

(21) Better estimates are available. For the first inequality see Szegö [5, Theorem 8.91.3]. The second inequality follows from the first combined with formula (2.9.6) below. For the properties of Hermite polynomials used in this discussion see also §§5.5, 5.7, 6.31, and 6.32 of Szegö's treatise.

$|x|\to\infty$. It is obvious that all $L$-transforms also belong to $C_0^{(0)}[-\infty,\ \infty]$. But much more can be asserted.

To this end we note that $f(x)\in F$ implies $x\,f(x)$ and $f'(x)\in F$. This is a consequence of the relations

$$\left.\begin{matrix} xu_n(x) \\ u_n'(x)\end{matrix}\right\} = -2^{-1/2}\{n^{1/2}u_{n-1}(x) \pm (n+1)^{1/2}u_{n+1}(x)\}, \tag{2.9.6}$$

which in their turn follow from the recurrence formulas for Hermite polynomials plus the relation $H_n'(x)=2n\,H_{n-1}(x)$. Hence

$$\left.\begin{matrix} xf(x) \\ f'(x)\end{matrix}\right\} = -2^{-1/2}\sum_{n=0}^{\infty}[(n+1)^{1/2}f_{n+1} \pm n^{1/2}f_{n-1}]u_n(x),$$

and these series are clearly elements of $F$. By induction we show that $x^mf^{(k)}(x)\in F$ for every $k$ and $m$. But this implies $x^mf^{(k)}(x)\in C_0^{(0)}[-\infty,\ \infty]$ for all $k$ and $m$.

Conversely, suppose that $x^mf^{(k)}(x)\in C_0^{(0)}[-\infty,\ \infty]$ for all $k$ and $m$. This implies that $x^mL^kf(x)$ has the same property and consequently $L^kf(x)\in L_2(-\infty,\ \infty)$ for every $k$. But if $g(x)$ is a function satisfying the same conditions as $f(x)$, an application of Lagrange's identity combined with the boundary conditions gives

$$\int_{-\infty}^{\infty}g(x)L^kf(x)dx = \int_{-\infty}^{\infty}f(x)L^kg(x)dx,$$

and in particular

$$\int_{-\infty}^{\infty}u_n(x)L^kf(x)dx = \int_{-\infty}^{\infty}f(x)L^ku_n(x)dx = (-1)^k(2n+1)^kf_n.$$

We conclude that the coefficients $f_n$ must satisfy conditions $C_1$ and $C_2$ and that $f(x)\in F$. Consequently we have proved:

*The class $F\{D^2-x^2,\ u_n(x),\ 2n+1;\ (-\infty,\ \infty)\}$ is that subset of $C^{(\infty)}[-\infty,\ \infty]$ the elements of which satisfy the boundary conditions*

$$\lim_{|x|\to\infty} x^mf^{(k)}(x) = 0, \qquad k, m = 0, 1, 2, \cdots. \tag{2.9.7}$$

From this we get without difficulty([22]):

*The class $F\{D^2-2x\,D,\ A_nH_n(x),\ 2n;\ (-\infty,\ \infty)\}$ is that subset of $C^{(\infty)}(-\infty,\ \infty)$, the elements of which satisfy the boundary conditions*

$$\lim_{|x|\to\infty} x^m\exp[-x^2/2]f^{(k)}(x) = 0, \quad k, m = 0, 1, 2, \cdots. \tag{2.9.8}$$

We can consequently formulate Theorem 9 as follows for the case of the Hermite operator.

([22]) $A_n$ is a normalization factor.

THEOREM 13. *Let $L=D^2-2x\,D$. Let $f(x)\in C^{(\infty)}(-\infty,\infty)$ and satisfy the boundary conditions* (2.9.8). *If* $\liminf_{k\to\infty} V[L^kf(x)]=N$, *then $f(x)$ is a polynomial of degree $N$. Conversely, every real polynomial of exact degree $N$ has the property $V[L^kf(x)]=N$ for all large values of $k$.*

2.10. **The Laguerre operator.** Our last example of a singular operator is that of Laguerre

$$L = xD^2 + (1-x)D, \tag{2.10.1}$$

the interval being $(0,\infty)$. The equation

$$(L-\lambda)y \equiv xy'' + (1-x)y' - \lambda y = 0$$

has singular points at 0 and $\infty$. A simple computation shows that $R(x, 1; \lambda) \to \infty$ at both points, so the problem is of type $T_3$. The origin is a regular singular point with indicial equation $\rho^2=0$. Hence there is a solution which becomes infinite as $\log(1/x)$ while the other solution is regular at $x=0$. The point at infinity is irregular-singular. Assuming $x$ and $\lambda$ positive, we have one solution tending to zero as $x^{-\lambda}$ and another tending to infinity as $x^{\lambda-1}e^x$ when $x\to\infty$. It follows that

$$y(x, 1; \lambda) = \begin{cases} A(\lambda)\log(1/x)\{1+o(1)\}, & x\to 0, \\ B(\lambda)x^{\lambda-1}e^x\{1+o(1)\}, & x\to\infty. \end{cases}$$

The singular boundary value problem

$$(L+\mu)u = 0, \qquad \lim_{x\to 0}\frac{u(x)}{\log(1/x)} = 0, \qquad \lim_{x\to\infty} u(x)x^{1-\lambda}e^{-x} = 0 \tag{2.10.2}$$

(for every $\lambda>0$) determines the Laguerre polynomials $L_n(x)$ corresponding to the characteristic values $\mu_n=n$, $n=0, 1, 2, \cdots$ [22].

The system $\{e^{-x/2}L_n(x)\}$ is complete in $L_2(0,\infty)$. It was proved by Szegö that

$$e^{-x/2}|L_n(x)| < 1, \qquad x > 0.$$

Since

$$L_n'(x) = -\sum_{\nu=0}^{n-1} L_\nu(x), \tag{2.10.3}$$

[22] To prove that no other solutions exist is fairly complicated. We shall merely outline an argument which the interested reader will be able to complete. There exist two formal but asymptotic solutions of the form $x^{\mu}\mathfrak{P}_1(1/x)$ and $e^x x^{-1-\mu}\mathfrak{P}_2(1/x)$, of which only the first one satisfies the boundary condition at infinity. The series are easily computed. If $\mu=n$ the first series terminates and reduces to a multiple of $L_n(x)$. For other values of $\mu$ it may be summed by Borel's method which leads to the result $u(x)=x^{\mu+1}\int_0^\infty F(-\mu, -\mu, 1, -t)e^{-xt}dt$. The behavior of the integral for small positive $x$ is determined by that of the hypergeometric function for large $t$. If $\mu$ is not zero or a positive integer, $F(-\mu, -\mu, 1, -t)=A(\mu)t^{\mu}\log t[1+o(1)]$, $A(\mu)\neq 0$, for large $t$, and $u(x)$ becomes logarithmically infinite when $x\to 0$. Thus the Laguerre polynomials are the only solutions of the boundary value problem.—For the properties of $L_n(x)$ used in this section, see Szegö [5, §§5.1, 5.7, 6.31, and 7.21].

we get

$$e^{-x/2}|L_n'(x)| < n, \qquad x > 0.$$

These inequalities show that $A_5$ and $E_3$ are satisfied if we take $U(x)=e^{x/2}$. For later use we note the recurrence formula

$$(2.10.4) \qquad (n+1)L_{n+1}(x) + (x-2n-1)L_n(x) + nL_{n-1}(x) = 0.$$

The zeros of $L_n(x)$ are all real positive and the $m$th zero equals $C_{m,n}(m+1)^2(n+1)^{-1}$ where $1/4 \leqq C_{m,n} \leqq 4$. It follows that $A_6$ also holds and $S$ is consequently a conservative system.

It remains to determine the class $F$. If $f(x)\in F$ then

$$(2.10.5) \qquad f(x) = \sum_{n=0}^{\infty} f_n L_n(x), \qquad \sum_{n=1}^{\infty} n^m |f_n| < \infty$$

for all $m$. Multiplying on both sides in (2.10.5) by $e^{-x/2}$ we obtain a series which is uniformly convergent in $[0, \infty]$ and the terms of which tend to zero when $x\to\infty$. It follows that $e^{-x/2}f(x)$ is continuous in $[0, \infty]$ and tends to zero when $x\to\infty$. We show next that $x f(x)$ and $f'(x)$ must belong to $F$ whenever $f(x)$ does. Multiplying both sides of (2.10.5) by $x$, reducing with the aid of (2.10.4) and rearranging, we obtain the series

$$xf(x) = \sum_{n=0}^{\infty} [(2n+1)f_n - nf_{n-1} - (n+1)f_{n+1}]L_n(x)$$

which clearly belongs to $F$. Similarly we obtain the series

$$f'(x) = -\sum_{n=0}^{\infty}\left(\sum_{\nu=n+1}^{\infty} f_\nu\right)L_n(x)$$

from (2.10.5) with the aid of (2.10.3). This is also an element of $F$. It follows that any function of the form $x^m f^{(k)}(x)\in F$ and

$$(2.10.6) \qquad \lim_{x\to\infty} x^m e^{-x/2} f^{(k)}(x) = 0, \qquad k, m = 0, 1, 2, \cdots.$$

At the origin we find of course that $f^{(k)}(x)$ tends to a finite limit for every $k$.

Conversely, if $f(x)\in C^{(\infty)}[0, \infty)$ and satisfies the boundary conditions (2.10.6), then $e^{-x/2}L^k f(x)\in L_2(0, \infty)$ for every $k$. Using Lagrange's identity we verify that

$$\int_0^\infty e^{-x/2}L_n(x)L^k f(x)dx = (-n)^k f_n$$

and from this we conclude that $f(x)\in F$.

*The class* $F\{xD^2+(1-x)D, L_n(x), n; (0, \infty)\}$ *equals the subset of* $C^{(\infty)}[0, \infty)$ *the elements of which satisfy conditions* (2.10.6).

THEOREM 14. *Let* $L=x D^2+(1-x) D$. *Let* $f(x)\in C^{(\infty)}[0, \infty)$ *and satisfy the*

*boundary conditions* (2.10.6). *If* $\liminf_{k\to\infty} V[L^k f(x)]=N$, *then* $f(x)$ *is a polynomial of degree* $N$. *Conversely, if* $f(x)$ *is a real polynomial of exact degree* $N$, *then* $V[L^k f(x)]=N$ *for all large values of* $k$([24]).

Theorems 12, 13, and 14 give three distinct unique characterizations of real ordinary polynomials in terms of their behavior with respect to certain second order differential operators. This is analogous to the unique characterization of trigonometric polynomials by means of the operator $D^2$ given by Pólya and Wiener.

2.11. **Bessel operators.** Our last examples will deal with semi-singular operator problems related to the theory of Bessel functions. In this theory we find essentially three different types of expansions, conventionally referred to as the Bessel-Fourier, the Neumann, and the Schlömilch series. Only the first type falls directly under our theory, but the third type is also accessible to the methods of Pólya and Wiener.

We start with the operator

$$L = D^2 + D/x - m^2/x^2 \tag{2.11.1}$$

where $m \geqq 0$ is fixed. We take the interval (0, 1) of which one end point is singular and the other regular. It is easily seen that $R(x, 1; \lambda)\to\infty$ when $x\to 0$ so the problem is of type $T_4$. We have $y(x; 1, \lambda)\sim B(\lambda)\log(1/x)$ or $B(\lambda)x^{-m}$ at $x=0$ according as $m=0$ or is greater than 0. The corresponding boundary value problem for $m>0$ is

$$(L+\mu)u = 0, \qquad \lim_{x\to 0} x^m u(x) = 0, \qquad C_1u(1) + C_2u'(1) = 0, \tag{2.11.2}$$

where $C_1 \geqq 0$, $C_2 \geqq 0$, $C_1+C_2>0$. If $m=0$ the factor $x^m$ should be replaced by $[\log(1/x)]^{-1}$. The problem has as its solution the set $J_m(\mu_n x)$, where $\mu_n$ runs through the positive roots of the equation

$$C_1\mu J_m(\mu) + C_2J_m(\mu) = 0. \tag{2.11.3}$$

If $m=0$, $C_2=0$, we have to add $\mu_0=0$ with $u_0(x)=1$([25]).

Using any standard text on Bessel functions, the reader will have no difficulties in proving that the corresponding system $S$ is conservative. We note in particular that $Z_n(0, 1) = V[J_m(\mu_n x)]=n$ so that $M(N)=N$ in Theorem 9. We shall not state the corresponding form of the theorem, but we shall determine the class $F$.

If $f(x)\in F\subset B_4^{(\infty)}\{L; C_1, C_2; (0, 1]\}$ then we must have

$$C_1L^nf(1) + C_2DL^nf(1) = 0, \qquad n = 0, 1, 2, \cdots, \tag{2.11.4}$$

([24]) Similar results hold for the general Laguerre operator $L=xD^2+(1+\alpha-x)D$, $\alpha>-1$.

([25]) The only solution which satisfies the boundary condition at $x=0$ is a multiple of $J_m(\mu x)$. The values of $\mu$ are determined by the second condition. The root $\mu=0$ figures if and only if $C_2/C_1=-m$. Owing to our sign restrictions, this case occurs only if $m=0$, $C_2=0$.

with our usual notation. At the singular end point we can write $f(x)=x^m g(x)$. A simple computation shows that if $g(x)$ is defined in $[-1, +1]$ by the convention $g(-x)=g(x)$ then $g(x)\in C^{(\infty)}[-1, +1]$.

Suppose, conversely, that $f(x) = x^m g(x)$ where $g(-x) = g(x)$, $g(x) \in C^{(\infty)}[-1, +1]$, and (2.11.4) is satisfied, The computation shows that $Lf(x)$ satisfies the same conditions. We can then apply the identity of Lagrange and find that

$$\int_0^1 xu_n(x)L^k f(x)dx = \int_0^1 xf(x)L^k u_n(x)dx = (-\mu_n)^k f_n,$$

since all intermediary integrated expressions vanish at both end points of (0, 1). It follows that the coefficients $f_n$ satisfy the conditions $C_1$ and $C_2$ of §2.1 and $f(x)\in F$.

Thus, *the class* $F\{L, A_n J_m(\mu_n x), \mu_n; (0, 1)\}$ *consists of all functions of the form* $f(x)=x^m g(x)$, *satisfying* (2.11.4), *such that* $g(-x)=g(x)$ *and* $g(x) \in C^{(\infty)}[-1, +1]$.

The Neumann series

$$\sum_{n=0}^{\infty} a_n J_n(x) \tag{2.11.5}$$

does not give rise to any interesting oscillation problems for the simple reason that in any fixed interval $(0, b)$ the function $J_n(x)$ is ultimately non-oscillatory since the least positive zero of $J_n(x)$ exceeds $n$.

We get more interesting results for the Schlömilch series

$$(f_0/2) + \sum_{n=1}^{\infty} f_n J_0(nx) \tag{2.11.6}$$

the terms of which are characteristic functions of the operator (2.11.1) with $m=0$ corresponding to the characteristic values $n^2$. The corresponding system $S$ is not admissible in the technical sense of §2.1, since the functions $\{J_0(nx)\}$ do not form an orthogonal system. But the methods employed in the present paper nevertheless apply and lead to a result which we state without proof[26].

THEOREM 15. *Let F be the class of functions defined by the formula*

$$f(x) = (2/\pi)\int_0^{\pi/2} g(x \sin t)dt, \qquad g(u) = (f_0/2) + \sum_{n=1}^{\infty} f_n \cos nu, \tag{2.11.7}$$

*where* $g(u)$ *is any real even function of period* $2\pi$ *belonging to* $C^{(\infty)}(-\infty, \infty)$. *Let* $N_k(R)$ *be the number of sign changes of* $\{D^2+(1/x)\, D\}^k f(x)$ *in the interval*

[26] See E. T. Whittaker and G. N. Watson [6, §17.82], for the relation between the series (2.11.6) and (2.11.7).

$(0, R)$. *If*

$$\liminf_{k\to\infty} \limsup_{R\to\infty} N_k(R)/R = C < \infty,$$

*then* $g(u)$ *is a trigonometric polynomial of degree at most* $C\pi$.

## Appendix

3.1. **Characteristic series representing entire functions.** Pólya and Wiener [2, Theorem III], proved for periodic functions $f(x)$ that the assumption $V[D^{2k}f(x)] = o(k^{1/2})$ implies that $f(x)$ is an entire function. This result also admits of far reaching generalizations but cannot be true for arbitrary conservative systems. It is obviously necessary that the functions $u_n(x)$ themselves are entire. It is also necessary to have some definite information concerning the convergency properties of the series $\sum f_n u_n(z)$ for complex values. In this direction it is enough to know that a condition of the form

$$\limsup_{n\to\infty} |f_n| \exp(\tau\mu_n^{1/2}) = \infty$$

for some finite $\tau$, prevents the convergence of the series in the whole finite plane, while on the other hand the finiteness of the limit superior for every $\tau$ implies that the series does converge in the whole plane. The matter is complicated by the fact that an entire function may have a characteristic series which is not convergent outside of the real interval $(a, b)$ or even anywhere([27]). In addition it is desirable to have more precise information concerning the characteristic values and the degree of regularity of the oscillations of the characteristic functions in fixed interior intervals. It is not worth while stating here in precise form the assumptions under which we have succeeded in extending Theorem III of Pólya and Wiener. It is enough to mention that the results apply to the operators of Legendre, Jacobi, Hermite and Laguerre. We state without proof:

Theorem 16. *The condition* $V[L^k f(x)] = o(k^{1/2})$ *is sufficient in order that* $f(x) = \sum_{n=0}^{\infty} f_n u_n(x)$ *shall define an entire function, the series being convergent in the finite complex plane, provided* $S = S\{L, u_n(x), \mu_n; (a, b)\}$ *is one of the five systems considered in* §§2.7 *to* 2.10.

For the case of the Legendre operator, this theorem has also been proved by Szegö ([4], special case of his Theorem D) and with a much less restrictive condition on the rate of growth of $N_k$. His method would also apply to the Jacobi case, at least for $\alpha > -1$, $\beta > -1$, with a similar improvement of the rate of growth condition. His method, however, does not apply to the Hermite, Hermite-Weber, and Laguerre operators.

([27]) This happens, for instance, in the case of expansions in terms of Hermite and Laguerre polynomials, but not in the Jacobi and Legendre cases.

3.2. **Upper limits for the frequency of oscillation.** It has been conjectured (by Pólya, at least for the operator $D$) that $o(k)$ is the correct order in theorems of the type of our Theorem 16 and that this order cannot be raised to $O(k)$. The latter part of the conjecture has been proved by Pólya and Szegö [4, §7]. It is very easy to verify that $O(k)$ is not admissible in the case of two rather wide classes of second order operators.

Suppose first that $(a, b)$ is a finite or semi-infinite interval and that the coefficients $p_m(x)$ of $L$ are polynomials. Take $f(x)=1/(x-c)$, where $c$ is real and outside of $[a, b]$. A simple computation shows that the $L^k$-transform of $f(x)$ is a rational function whose denominator is $(x-c)^{2k+1}$ while the numerator is a polynomial of degree at most $Ak$, where $A$ is a constant depending only upon the degree of the polynomials $p_m(x)$. It is clear that for this function $V[L^kf(x)]\leqq Ak$ and $f(x)$ is not entire. If $(a, b)=(-\infty, \infty)$, we take $f(x)=1/(x^2+c^2)$ instead.

If $(a, b)=(-\pi, \pi)$ and the coefficients $p_m(x)$ are trigonometric polynomials, we have similar results. We take $f(x)=1/(2-\sin x)$ instead. Here $L^kf(x)$ is the quotient of two trigonometric polynomials, the degree of the numerator being at most $Ak$. Hence $V[L^kf(x)]\leqq 2Ak$ and $f(x)$ is not entire.

Finally it should be observed that all the available evidence so far supports the conjecture that $V[L^kf(x)]=O(k)$ is a necessary and sufficient condition in order that an admissible characteristic series shall define an analytic function.

## References


1. E. L. Ince, *Ordinary Differential Equations*, London, 1927.

2. G. Pólya and N. Wiener, *On the oscillation of the derivatives of a periodic function*, these Transactions, vol. 52 (1942), pp. 249–256.

3. L. Schlesinger, *Handbuch der Theorie der linearen Differentialgleichungen*, Leipzig, vol. I, 1895, vol. $II_1$, 1897, vol. $II_2$, 1898.

4. G. Szegö, *On the oscillation of differential transforms*. I, these Transactions, vol. 52 (1942), pp. 450–462.

5. ———, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, New York, vol. 23, 1939.

6. E. T. Whittaker and G. N. Watson, *A Course in Modern Analysis*, 4th edition, Cambridge, 1935.



Stanford University,
Stanford University, Calif.
Yale University,
New Haven, Conn.

# INTEGRATION IN A CONVEX LINEAR TOPOLOGICAL SPACE

BY

C. E. RICKART

## Introduction

The results of this paper center around the definition of an integration process for multi-valued set functions which are defined over a $\sigma$-field $\mathfrak{M}$ and whose values lie in a convex linear topological space $\mathfrak{X}$. As such they represent a substantial generalization of the basic results contained in a paper by A. Kolmogoroff [7](1), who considered the case in which $\mathfrak{X}$ is the real numbers. On the other hand, the method of defining the integral is a generalization of that used by R. S. Phillips(2) [12, p. 118], although Phillips considered integration only with respect to a positive numerical measure function and restricted the integral to be single-valued. The importance of the Phillips definition lies in the fact that it relieves one of the necessity of considering infinite sums. Throughout the paper is emphasized a type of convergence for sets in a linear topological space which is analogous to the Hausdorff notion of convergence for sets in a metric space [5, §28]. G. B. Price has made a similar use of the Hausdorff convergence for sets [13, Parts II, V].

The contents of the paper are divided into four parts. Part I (§§1–3) contains a short discussion of convex linear topological spaces, a definition of the notion of unconditional summability, which plays a central role in the definition of the integral, and two theorems on additive set functions. Part II (§§4–9) contains the general theory of the $\mathcal{V}$- and $\mathcal{S}\mathcal{V}$-integrals. The $\mathcal{V}$-integral is multi-valued and is defined for multi-valued set functions $F(\sigma)$. The $\mathcal{S}\mathcal{V}$-integral is the single-valued specialization of the $\mathcal{V}$-integral. Definitions and basic properties of these integrals account for §§4, 5. Section 6 contains a discussion of a generalization of the Kolmogoroff [7] notion of differential equivalence applied to the $\mathcal{V}$-integral, and §7 contains a proof that the transform of an integrable function by a general type of linear transformation is integrable. In §8 it is shown that the definition of the $\mathcal{S}\mathcal{V}$-integral can be weakened in case $\mathfrak{X}$ is complete in a certain sense. Section 9 contains a convergence theorem for the $\mathcal{S}\mathcal{V}$-integral which involves a generalization of the notion of approximate convergence to functions $F(\sigma)$ of the type considered here. The approximate convergence is relative to a positive numerical measure function $m(\sigma)$ whose only relation to the integral lies in the condition that

Presented to the Society, April 11, 1941; received by the editors September 8, 1941, and, in revised form, February 11, 1942.

(1) Numbers in brackets refer to the bibliography at the end of the paper.

(2) See also a paper by Garrett Birkhoff [2, p. 51] where the same definition used by Phillips is given.

the integrals of the functions considered be absolutely continuous relative to $m(\sigma)$. Part III (§§10–13) is concerned with the $\mathcal{V}_B$-integral which is a specialization of the $\mathfrak{S}\mathcal{V}$-integral and may be described as integration with respect to a "bilinear" function. The "bilinear" function $B[y, \sigma]$ is a generalization of $m(\sigma)y$, where $m(\sigma)$ is a numerical measure function. It has its values in $\mathfrak{X}$ and is defined for $y$ in a linear space $\mathfrak{Y}$ and $\sigma$ in the $\sigma$-field $\mathfrak{M}$. It is linear in $y$ for each $\sigma$ and completely additive in $\sigma$ for each $y$. Functions $y(\sigma)$ to be integrated have their values in $\mathfrak{Y}$ while the value of the integral is in $\mathfrak{X}$. Section 10 contains a definition of the $\mathcal{V}_B$-integral and a discussion of its fundamental properties. In §11 the $\mathcal{V}_B$-integral is considered for the case in which $\mathfrak{X}$ is complete in the sense of §8. Section 12 contains a discussion of absolute continuity and a convergence theorem for the $\mathcal{V}_B$-integral. In §13 the existence of the $\mathcal{V}_B$-integral is proved for a certain class of measurable functions. Part IV (§§14–16) relates the above integrals to previously defined integrals. For the case in which $\mathfrak{X}$ is the real numbers, the $\mathfrak{S}\mathcal{V}$-integral includes an integral of Kolmogoroff [7]. The $\mathcal{V}_B$-integral reduces in a special case to an integral of Phillips [12], and a specialization of the $\mathcal{V}_B$-integral includes an integral of Price [13]. Relation of the $\mathcal{V}_B$-integral to the various other integrals which have been defined can be obtained through its relation to the Phillips integral (see [12, §7]).

## Part I. Preliminary considerations

1. **Convex linear topological spaces.** The type of linear topological space $\mathfrak{X}$ to be considered here is that introduced by J. von Neumann [11, p. 4]. It is defined as follows:

A set $\mathfrak{X}$ of elements $x$ is said to constitute a *linear topological space* provided it is linear([3]) (Banach [1, p. 26]) and provided it contains a family $\mathcal{V}$ of subsets such that

(1) $\theta \in V$ for every $V \in \mathcal{V}$.

(2) $x \in V$ for every $V \in \mathcal{V}$ implies $x = \theta$.

(3) $V_1, V_2 \in \mathcal{V}$ implies the existence of $V_3 \in \mathcal{V}$ such that([4]) $V_3 \subset V_1 \cap V_2$.

(4) $V \in \mathcal{V}$ implies the existence of $V' \in \mathcal{V}$ such that([5]) $V' + V' \subset V$.

(5) $V \in \mathcal{V}$ implies the existence of $V' \in \mathcal{V}$ such that(I) $\alpha V' \subset V$ for all $|\alpha| \leqq 1$.

(6) $x \in \mathfrak{X}$ and $V \in \mathcal{V}$ imply the existence of $\alpha$ such that $x \in \alpha V$.

Also, $\mathfrak{X}$ is said to be *convex* provided

(7) $V \in \mathcal{V}$ implies $V + V \subset 2V$.

---

([3]) Scalar multipliers are assumed real. The zero element will be denoted by $\theta$.

([4]) The symbols $\subset$, $\cap$, $\cup$ denote, respectively, set-theoretic "included in," "intersection" and "union," and will be used with their usual variations throughout the paper. $A \cap CB$ denotes the set of points contained in $A$ but not in $B$.

([5]) $V_1 + V_2 = \{x_1 + x_2 \mid x_1 \in V_1, x_2 \in V_2\}$, where $\{x \mid P\}$ denotes the set of all elements $x$ subject to the condition $P$. Similarly, $\alpha V = \{\alpha x \mid x \in V\}$.

A set $G$ in $\mathfrak{X}$ is defined to be *open* provided, for every $x \in G$, there exists $V \in \mathcal{V}$ such that $x + V \subset G$. The *interior* of a set $X$ is defined by $X_i = \{x \mid x + V \subset X \text{ for some } V \in \mathcal{V}\}$. $X_i$ is evidently open. A set is defined to be *closed* if it is the complement of an open set, and the *closure* of a set $X$ is defined by $X_{cl} = C((CX)_i)$. Evidently $X_{cl}$ is closed. The above class of open sets defines a regular Hausdorff topology in $\mathfrak{X}$ so that the operations of addition and multiplication by a number are continuous [11, Theorem 6]. It is known (Wehausen [16, Theorem 1]) that this topology is equivalent to that introduced by Kolmogoroff [8, p. 29]. In all that follows $\mathfrak{X}$ will be assumed to be a convex linear topological space as above defined. An important consequence of the convexity of the space $\mathfrak{X}$ is that the closure $V_{cl}$ of every $V \in \mathcal{V}$ is a convex set; that is, $C_o V_{cl} = V_{cl}$, where $C_o X = \{\sum_{i=1}^{n} \alpha_i x_i \mid x_i \in X, \alpha_i \geqq 0, \sum_{i=1}^{n} \alpha_i = 1, n \text{ arbitrary}\}$. This implies that $\sum_{i=1}^{n} \alpha_i \mathcal{V}_{cl} = (\sum_{i=1}^{n} \alpha_i) V_{cl}$ for arbitrary $\alpha_i \geqq 0$. Another important consequence of the convexity of $\mathfrak{X}$ is that $0 < \alpha < \beta$ implies $\alpha V_{cl} \subset \beta V$ for every $V \in \mathcal{V}$. In much of the material which follows, the assumption that $\mathfrak{X}$ is convex could be avoided; however computation is greatly simplified by using it and some of the important theorems (for example, Theorems 3.1, 5.5, 9.5) seem to involve it rather deeply.

It will be desirable later(6) to subject $\mathfrak{X}$ to a completeness condition which we introduce here. It is convenient to define the condition in terms of the Moore-Smith [10, p. 103] general limits notion on a class $\mathcal{L}$ with a transitive compositive relation $R$ on $\mathcal{L}\mathcal{L}$. A set $\{x_l\}$ of elements in $\mathfrak{X}$, where $l$ ranges over $\mathcal{L}$, is called an *$\mathcal{L}$-directed* set. $\{x_l\}$ is called a *fundamental* $\mathcal{L}$-directed set provided, for every $V \in \mathcal{V}$, there exists $l_V$ such that $l_i R l_V$ $(i = 1, 2)$ implies $x_{l_1} - x_{l_2} \in V$. The space $\mathfrak{X}$ is said to be *complete relative to* $\mathcal{L}$ provided every fundamental $\mathcal{L}$-directed set converges (in the Moore-Smith sense) to an element of $\mathfrak{X}$. We will be interested in two important specializations of this completeness notion, for example, the case where $\mathcal{L}$ is the set of positive integers and $R$ is the usual order relation "$>$," which gives the ordinary sequential completeness, and the case where $\mathcal{L}$ is the family of neighborhoods $\mathcal{V}$ and $R$ is the set-theoretic "included in" (that is, $V_1 R V_2$ means $V_1 \subset V_2$)(7). It is easy to prove that, if $\mathfrak{X}$ satisfies the first countability axiom (Hausdorff [5, p. 229]) and is sequentially complete, then it is complete relative to $\mathcal{V}$. It follows that, if $\mathfrak{X}$ is a Banach space with its norm topology (that is, $\mathcal{V}$ is the family of spheres with center $\theta$), then $\mathfrak{X}$ is complete relative to $\mathcal{V}$.

2. **Unconditional summability.** In the following, $\pi$ will always denote a finite set of positive integers and $\pi_1 \geqq \pi_2$ will mean that $\pi_1$ contains $\pi_2$. Also, $\sum_\pi$ will mean summation over those $n \in \pi$. Two subsets $X$, $Y$ of $\mathfrak{X}$ are said to be *equal within* $V$ provided $X \subset Y + V$ and $Y \subset X + V$.

(6) See Theorems 8.5, 12.3 below.

(7) The space $\mathfrak{X}$ may be said to be *complete* provided it is complete relative to *every* $\mathcal{L}$. This condition can be shown to be equivalent to simply completeness relative to $\mathcal{V}$ (see Graves [4, p. 62]).

2.1. DEFINITION. *Two sequences* $\{X_n\}$, $\{X_n'\}$ *of subsets of* $\mathfrak{X}$ *are said to be summably equal within* $V$ *provided there exists a* $\pi_0$ *such that* $\pi \geqq \pi_0$ *implies* $\sum_\pi X_n$, $\sum_\pi X_n'$ *are equal within* $V$.

2.2. DEFINITION. *A sequence* $\{X_n\}$ *of subsets of* $\mathfrak{X}$ *is said to be unconditionally summable (we write u.s.) to a set* $X$ *with respect to* $V$ *provided there exists* $\pi_0$ *such that* $\pi \geqq \pi_0$ *implies* $X$, $\sum_\pi X_n$ *are equal within* $V$.

Observe(8) that $\{X_n\}$ is u.s. to a single element $x \in \mathfrak{X}$ with respect to $V$ provided there exists $\pi_0$ such that $\pi \geqq \pi_0$ implies $\pm \{\sum_\pi X_n - x\} \subset V$. This special case of the above notion of unconditional summability was used by R. S. Phillips [12, p. 118]. Observe also that, if each of the sets $X_n$, $X$ consists of only a single element, then u.s. of $\{X_n\}$ to $X$ with respect to every $V \in \mathcal{V}$ reduces to the ordinary unconditional convergence.

2.3. THEOREM. *In order that* $\{X_n\}$ *be u.s. to* $X$ *with respect to* $V$ *it is both necessary and sufficient that, for every rearrangement* $n(k)$ *of the sets* $X_n$, *there exists* $k_0$ *for which* $k \geqq k_0$ *implies* $X$, $\sum_{i=1}^{k} X_{n(i)}$ *are equal within* $V$.

The method of proof used in an analogous situation involving unconditional convergence of series in normed vector spaces can be applied with slight modification here (see Hildebrandt [6, p. 90]).

3. **Additive set functions.** Let $M$ be an abstract set of elements and $\mathfrak{M}$ a $\sigma$-field(9) of subsets of $M$. Elements of $\mathfrak{M}$ will be denoted by $\sigma$. Also let $m(\sigma)$ be a positive, completely additive measure function defined over $\mathfrak{M}$. A single-valued function $x(\sigma)$ on $\mathfrak{M}$ to the space $\mathfrak{X}$ is said to be *completely additive* if for every sequence $\{\sigma_n\}$ of disjoint elements of $\mathfrak{M}$, the series $\sum x(\sigma_n)$ is unconditionally convergent to $x(\cup \sigma_n)$. $x(\sigma)$ is said to be *absolutely continuous relative to* $m(\sigma)$ provided, for every $V \in \mathcal{V}$, there exists $\delta_V > 0$ such that $x(\sigma) \in V$ whenever $m(\sigma) < \delta_V$.

3.1. THEOREM(10). *If* $x(\sigma)$ *is completely additive on* $\mathfrak{M}$ *to* $\mathfrak{X}$ *and if* $m(\sigma) = 0$ *implies* $x(\sigma) = \theta$, *then* $x(\sigma)$ *is absolutely continuous relative to* $m(\sigma)$.

Suppose(11) the theorem not true; then there exists a $V \in \mathcal{V}$ and a sequence of elements $\sigma_k \in \mathfrak{M}$ such that $\lim_{n\to\infty} m(\cup_{k=n}^{\infty} \sigma_k) = 0$ and such that(12) $\|x(\sigma_k)\|_V > 2$ for each $k$. Now, by a result of Wehausen [16, Theorem 8], there exists a linear continuous operation $\bar{x}_1$ on $\mathfrak{X}$ to the real numbers such that $|\bar{x}_1(x)| \leqq \|x\|_V$ for all $x$ and $\bar{x}_1(x(\sigma_1)) = \|x(\sigma_1)\|_V$. It is obvious that

(8) Unconditional summability obviously involves a convergence notion related to the well known Hausdorff convergence for sets in a metric space [5, §28].

(9) The $\sigma$-field $\mathfrak{M}$ has the following properties: (1) $\mathfrak{M}$ contains the empty set; (2) if $\sigma \in \mathfrak{M}$, then $M \cap C\sigma \in \mathfrak{M}$; (3) if $\sigma_n \in \mathfrak{M}$ $(n = 1, 2, \cdots)$, then $\cup_{n=1}^{\infty} \sigma_n \in \mathfrak{M}$.

(10) See Dunford [3, Theorem 42] and Kunisawa [9, pp. 68, 69].

(11) The method of proof used here was suggested to the writer by R. S. Phillips.

(12) $\|x\|_V$ is the von Neumann pseudo-norm; that is, $\|x\|_V = \max(\|x\|_V^+, \|-x\|_V^+)$, where $\|x\|_V^+ = \text{g.l.b. } \{\alpha \mid \alpha > 0, x \in \alpha V\}$ [11, pp. 18, 19].

$f(\sigma)=\bar{x}_1(x(\sigma))$ is a completely additive, real-valued function of $\sigma$ such that $m(\sigma)=0$ implies $f(\sigma)=0$. Therefore, by a well known theorem (see Saks [15, Theorem 13.2, p. 31]), $f(\sigma)$ is absolutely continuous relative to $m(\sigma)$. It follows that there exists an $n_1$ such that $|f(\bigcup_{k=n}^{\infty}(\sigma_1\cap\sigma_k))|<1$ for $n\geqq n_1$. Now, if we take(13) $\sigma_1^0=\sigma_1\cap C[\bigcup_{k=n_1}^{\infty}(\sigma_1\cap\sigma_k)]$, it is clear that $f(\sigma_1)=f(\sigma_1^0)+f(\bigcup_{k=n_1}^{\infty}(\sigma_1\cap\sigma_k))$. But $f(\sigma_1)=\bar{x}_1(x(\sigma_1))=\|x(\sigma_1^0)\|_V>2$; therefore, $f(\sigma_1^0)=\bar{x}_1(x(\sigma_1^0))>1$. Since $|\bar{x}_1(x(\sigma_1^0))|\leqq\|x(\sigma_1^0)\|_V$, we have $\|x(\sigma_1^0)\|_V>1$. Moreover, $\sigma_1^0\cap\sigma_n=0$ for all $n\geqq n_1$. Repeating the above procedure on the sequence $\sigma_{n_1}, \sigma_{n_1+1}, \cdots$ one obtains a $\sigma_2^0\subset\sigma_{n_1}$ and an $n_2$ such that $\sigma_2^0\cap\sigma_n=0$ for $n\geqq n_2$ and $\|x(\sigma_2^0)\|_V>1$. This process can be continued indefinitely to obtain a sequence $\{\sigma_n\}$ of *disjoint* elements of $\mathfrak{M}$ such that $\|x(\sigma_n^0)\|_V>1$ for all $n$. But this result obviously contradicts the assumption that $x(\sigma)$ be completely additive; therefore $x(\sigma)$ is absolutely continuous relative to $m(\sigma)$.

3.2. THEOREM. *Let each of the functions $x_n(\sigma)$ be additive and absolutely continuous relative to $m(\sigma)$. Then, if $\{x_n(\sigma)\}$ is a fundamental sequence for each $\sigma$, the $x_n(\sigma)$ are equi-absolutely continuous relative to $m(\sigma)$.*

A proof indentical with that which gives Theorem 6.1 of the Phillips paper [12, p. 125] applies here; so will be omitted. The method of proof is due to Saks [14].

## PART II. THE GENERAL THEORY

4. **Definitions of the $\mathcal{V}$- and $S\mathcal{V}$-integrals.** Let $\mathfrak{M}$ denote, as before, a $\sigma$-field of subsets of an abstract set $M$. A subdivision of $M$ into a finite or denumerable number of disjoint elements of $\mathfrak{M}$ will be denoted by $\Delta=\{\sigma_i\}$, where $M=\bigcup\sigma_i$ and $\sigma_i\cap\sigma_j=0$ $(i\neq j)$. $\Delta^1$ is said to be *finer than* $\Delta^2$ provided, for every $\sigma_i^1$, there exists a $\sigma_{n_i}^2$ such that $\sigma_i^1\subseteq\sigma_{n_i}^2$; we write $\Delta^1\geqq\Delta^2$. The product of two subdivisions is a subdivision defined by $\Delta^1\Delta^2=\{\sigma_i^1\cap\sigma_j^2\}$. Evidently the product of two subdivisions is finer than either one of the subdivisions. Let $\Delta_0=\{\sigma_k\}$ be a given fixed subdivision and $\Delta^k$ $(k=1, 2, \cdots)$ an arbitrary sequence of subdivisions; then the subdivision $\Delta$ which coincides with $\Delta^k$ on the set $\sigma_k$ $(k=1, 2, \cdots)$ is called the *sum* of the $\Delta^k$ over $\Delta_0$.

The functions $F(\sigma)$ to be studied are multi-valued and are defined over(14) $\mathfrak{M}$ (that is, excluding the empty set) with values in $\mathfrak{X}$. Let $\Delta=\{\sigma_i\}$ be an arbitrary subdivision of $M$; we denote the sequence of sets $\{F(\sigma\cap\sigma_i)\}$ by the symbol $J(F, \sigma, \Delta)$.

4.1. DEFINITION. *The function $F(\sigma)$ is said to be $\mathcal{V}$-integrable over $\sigma_0$ provided there is a set $I(F, \sigma_0)\subset\mathfrak{X}$ such that, for every $V\in\mathcal{V}$, there exists $\Delta_{V\sigma_0}$ for which $\Delta\geqq\Delta_{V\sigma_0}$ implies $J(F, \sigma_0, \Delta)$ is u.s. to $I(F, \sigma_0)$ with respect to $V$. The*

(13) See Footnote 4 above.

(14) One could develop the following theory for functions defined over only a portion of $\mathfrak{M}$; however there would be a considerable loss of simplicity in the statement of definitions and theorems.

*closure*[15] *of the set* $I(F, \sigma_0)$ *will be called the* $\mathcal{V}$*-integral of* $F(\sigma)$ *over* $\sigma_0$, *and we write* $I(F, \sigma_0)_{cl} = \int_{\sigma_0} F(d\sigma)$. *Furthermore, if* $\int_{\sigma_0} F(d\sigma)$ *consists of a single element, then* $F(\sigma)$ *is said to be* $\mathfrak{S}\,\mathcal{V}$*-integrable over* $\sigma_0$.

4.2. THEOREM. *If* $F(\sigma)$ *is* $\mathcal{V}$*-integrable on* $\sigma_0$, *then* $\int_{\sigma_0} F(d\sigma)$ *is unique.*

Suppose $F(\sigma)$ $\mathcal{V}$-integrable to each of the sets $I_1(F, \sigma_0)$, $I_2(F, \sigma_0)$. Then it is immediate from the definition that, for every $V \in \mathcal{V}$, there exists a $\Delta_{V\sigma_0}$ such that $J(F, \sigma_0, \Delta)$ is u.s. to *both* $I_1(F, \sigma_0)$ and $I_2(F, \sigma_0)$ with respect to $V$ for $\Delta \geqq \Delta_{V\sigma_0}$; that is, if $\Delta = \{\sigma_i\}$, then there exists $\pi_0$ such that $\pi \geqq \pi_0$ implies $\sum_\pi F(\sigma_0 \cap \sigma_i)$, $I_m(F, \sigma_0)$ are equal within $V$ $(m = 1, 2)$. It follows immediately from this result that $I_1(F, \sigma_0)$, $I_2(F, \sigma_0)$ are equal within $2V$. Therefore, $I_k(F, \sigma_0) \subset I_l(F, \sigma_0) + 2V$ $(k = 1, 2;\ l = 1, 2)$. Since $V$ is arbitrary, $I_k(F, \sigma_0) \subseteq I_l(F, \sigma_0)_{cl}$; hence $I_1(F, \sigma_0)_{cl} = I_2(F, \sigma_0)_{cl}$.

Observe that, if $F(\sigma \cap \sigma_0) = \theta$ for every $\sigma$, then $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$ to the value $\theta$.

## 5. Properties of the integrals[16].

5.1. THEOREM. *If* $F(\sigma)$, $G(\sigma)$ *are* $\mathcal{V}$*-integrable on* $\sigma_0$ *and* $\alpha$ *is any real number, then* $\alpha F(\sigma)$, $F(\sigma) + G(\sigma)$ *are* $\mathcal{V}$*-integrable on* $\sigma_0$ *and*

$$\int_{\sigma_0} \alpha F(d\sigma) = \alpha \int_{\sigma_0} F(d\sigma), \qquad \int_{\sigma_0} F(d\sigma) + G(d\sigma) = \left[\int_{\sigma_0} F(d\sigma) + \int_{\sigma_0} G(d\sigma)\right]_{cl}.$$

If $\alpha = 0$, the statement for $\alpha F(\sigma)$ is obvious, and, if $\alpha \neq 0$, the desired result is a consequence of the fact that, if $J(F, \sigma_0, \Delta)$ is u.s. to $I(F, \sigma_0)$ with respect to $V$, then $J(\alpha F, \sigma_0, \Delta)$ is u.s. to $\alpha I(F, \sigma_0)$ with respect to $\alpha V$.

In the case of $F(\sigma) + G(\sigma)$, we observe that, for arbitrary $V \in \mathcal{V}$, there exists a $\Delta_{V\sigma_0}$ such that, if $\Delta \geqq \Delta_{V\sigma_0}$, then $J(F, \sigma_0, \Delta)$, $J(G, \sigma_0, \Delta)$ are, respectively, u.s. to $\int_{\sigma_0} F(d\sigma)$, $\int_{\sigma_0} G(d\sigma)$ with respect to $V$. From this it is immediate that $J(F+G, \sigma_0, \Delta)$ is u.s. to $\int_{\sigma_0} F(d\sigma) + \int_{\sigma_0} G(d\sigma)$ with respect to $2V$. Since $V$ is arbitrary, the desired result follows.

5.2. COROLLARY. *If* $F(\sigma)$, $G(\sigma)$ *are* $\mathfrak{S}\,\mathcal{V}$*-integrable on* $\sigma_0$, *then*

$$\int_{\sigma_0} \alpha F(d\sigma) = \alpha \int_{\sigma_0} F(d\sigma), \qquad \int_{\sigma_0} F(d\sigma) + G(d\sigma) = \int_{\sigma_0} F(d\sigma) + \int_{\sigma_0} G(d\sigma).$$

5.3. THEOREM. *If* $F(\sigma)$ *is* $\mathcal{V}$*-integrable on both* $\sigma_1$, $\sigma_2$ *and if* $\sigma_1 \cap \sigma_2 = 0$, *then* $F(\sigma)$ *is* $\mathcal{V}$*-integrable on* $\sigma_1 \cup \sigma_2$ *and*

---

(15) The closure of a set $X$ is denoted by $X_{cl}$. It can be shown (see [11]) that $X_{cl} = \bigcap(X + V)$ for $V \in \mathcal{V}$. Observe that, if $F(\sigma)$ is $\mathcal{V}$-integrable to $I(F, \sigma_0)$, then it is also $\mathcal{V}$ integrable to $I(F, \sigma_0)_{cl}$.

(16) The theorems of this section will be stated for the $\mathcal{V}$-integral and are, of course, true for the $\mathfrak{S}$ $\mathcal{V}$-integral. In case the results for the $\mathfrak{S}$ $\mathcal{V}$-integral are stronger, we state them as corollaries.

$$\int_{\sigma_1\cup\sigma_2} F(d\sigma) = \left[\int_{\sigma_1} F(d\sigma) + \int_{\sigma_2} F(d\sigma)\right]_{cl}.$$

5.4. COROLLARY. *If $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on both $\sigma_1$, $\sigma_2$ and if $\sigma_1\cap\sigma_2=0$, then $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_1\cup\sigma_2$ and*

$$\int_{\sigma_1\cup\sigma_2} F(d\sigma) = \int_{\sigma_1} F(d\sigma) + \int_{\sigma_2} F(d\sigma).$$

5.5. THEOREM. *The $\mathcal{V}$-integral is a completely additive function of $\sigma$ in the sense that, if $F(\sigma)$ is $\mathcal{V}$-integrable on each $\sigma^k$ ($k=0, 1, 2, \cdots$), where $\sigma^0=\bigcup\sigma^k$ and $\sigma^m\cap\sigma^n=0$ ($m\neq n$; $m, n\neq 0$), then $\{\int_{\sigma^k}F(d\sigma)\}$ ($k=1, 2, \cdots$) is u.s. to $\int_{\sigma^0}F(d\sigma)$ with respect to every $V\in\mathcal{V}$.*

There is no loss in taking $\sigma^0=M$. Since the integral exists for each $\sigma^k$, there exists $\Delta_V^k$ such that, if $\Delta\geqq\Delta_V^k$, $J(F, \sigma^k, \Delta)$ is u.s. to $\int_{\sigma^k}F(d\sigma)$ with respect to $2^{-k-1}V$ ($k=0, 1, 2, \cdots$). Denote by $\Delta_1$ the sum of the $\Delta_V^k$ ($k=1, 2, \cdots$) over the subdivision $\{\sigma^k\}$ (see §4 above) and set $\Delta_0=\Delta_1\Delta_V^0$. Evidently on the set $\sigma^k$ the subdivision $\Delta_0$ is finer than the subdivision $\Delta_V^k$. Therefore $J(F, \sigma^k, \Delta_0)$ is u.s. to $\int_{\sigma^k}F(d\sigma)$ with respect to $2^{-k-1}V$. If $\Delta_0=\{\sigma_i\}$, then there exists $\pi_0$ such that $\pi\geqq\pi_0$ implies the equality of $\sum_\pi F(\sigma_i)$, $\int_M F(d\sigma)$ within $V/2$. Set $n_V=\max\{n\mid\sigma^n\cap\bigcup_{\pi_0}\sigma_i\neq 0\}$; then, for an arbitrary (but fixed) $n\geqq n_V$, there exists a $\pi_n$ such that $\pi_n\geqq\pi_0$, $i\in\pi_n$, $l>n$ imply $\sigma^l\cap\sigma_i=0$ and

$$\sum_{i\in\pi_n} F(\sigma^k\cap\sigma_i)\subset\int_{\sigma^k}F(d\sigma)+2^{-k-1}V,\quad \int_{\sigma^k}F(d\sigma)\subset\sum_{i\in\pi_n}F(\sigma^k\cap\sigma_i)+2^{-k-1}V,$$

for $k=0, 1, 2, \cdots, n$. It is obvious that [17]

$$\sum_{k=1}^{n}\sum_{i\in\pi_n}F(\sigma^k\cap\sigma_i)=\sum_{\pi_n}F(\sigma_i);$$

therefore,

$$\sum_{\pi_n}F(\sigma_i)\subset\sum_{k=1}^{n}\int_{\sigma^k}F(d\sigma)+V/2,\qquad \sum_{k=1}^{n}\int_{\sigma^k}F(d\sigma)\subset\sum_{\pi_n}F(\sigma_i)+V/2.$$

Now, since $\pi_n\geqq\pi_0$ we have immediately that

$$\sum_{k=1}^{n}\int_{\sigma^k}F(d\sigma)\subset\int_M F(d\sigma)+V,\qquad \int_M F(d\sigma)\subset\sum_{k=1}^{n}\int_{\sigma^k}F(d\sigma)+V,$$

where $n\geqq n_V$ is arbitrary. This argument does not depend on the order in which the sets $\sigma^k$ are taken; therefore the desired result follows from Theorem 2.3.

[17] It is understood that terms for which $\sigma^k\cap\sigma_i=0$ are omitted.

5.6. COROLLARY. *The $\mathfrak{S}\mathcal{V}$-integral is completely additive in the ordinary sense.*

The proofs of the next two theorems are not difficult and will be omitted.

5.7. THEOREM. *Any function $F(\sigma)$ defined over $\mathfrak{M}$ which is completely additive in the sense of Theorem 5.5 is $\mathcal{V}$-integrable on every $\sigma$ and $F(\sigma)_{cl}=\int_\sigma F(d\sigma)$.*

5.8. THEOREM. *If $F(\sigma)$, $G(\sigma)$ are $\mathcal{V}$-integrable on $\sigma_0$ and if there exists $\Delta_0$ such that $\{\sigma_i\}=\Delta\geqq\Delta_0$ implies the existence of $\pi_\Delta$ such that, for $\pi\geqq\pi_\Delta$, it is true that*

$$\sum_\pi G(\sigma_0\cap\sigma_i)\subset\sum_\pi F(\sigma_0\cap\sigma_i)+V,$$

*then*

$$\int_{\sigma_0}G(d\sigma)\subseteq\left[\int_{\sigma_0}F(d\sigma)+V\right]_{cl}.$$

5.9. COROLLARY. *If $F(\sigma)$, $G(\sigma)$ satisfy the conditions of Theorem 5.8 and if in addition $F(\sigma)$ is $\mathfrak{S}\mathcal{V}$-integrable, then $\int_{\sigma_0}G(d\sigma)-\int_{\sigma_0}F(d\sigma)\subseteq V_{cl}$.*

5.10. THEOREM. *If $F(\sigma)$ is $\mathcal{V}$-integrable on $\sigma_0$, then so also is $F(\sigma)_{cl}$ and to the same value.*

For every $V\in\mathcal{V}$, there exists $\Delta_V$ such that, if $\{\sigma_i\}=\Delta\geqq\Delta_V$, then there exists $\pi_\Delta$ for which $\pi\geqq\pi_\Delta$ implies the equality of $\int_{\sigma_0}F(d\sigma)$, $\sum_\pi F(\sigma_0\cap\sigma_i)$ within $V$. Since $F(\sigma)_{cl}\subset F(\sigma)+V'$ for *arbitrary* $V'\in\mathcal{V}$ and $\sigma$, it follows that $\sum_\pi F(\sigma_0\cap\sigma_i)_{cl}\subset\sum_\pi F(\sigma_0\cap\sigma_i)+V$. Therefore $\int_{\sigma_0}F(d\sigma)$, $\sum_\pi F(\sigma_0\cap\sigma_i)_{cl}$ are equal within $2V$. Since $V$ is arbitrary, the desired result follows by definition.

5.11. THEOREM. *If $F(\sigma)$, $G(\sigma)$ are $\mathcal{V}$-integrable on $\sigma_0$ and if, for every $\sigma\subseteq\sigma_0$, $G(\sigma)\subseteq F(\sigma)_{cl}$, then $\int_{\sigma_0}G(d\sigma)\subseteq\int_{\sigma_0}F(d\sigma)$.*

In the present case the conditions of Theorem 5.8 hold for every $V\in\mathcal{V}$; therefore

$$\int_{\sigma_0}G(d\sigma)\subseteq\left[\int_{\sigma_0}F(d\sigma)+V/2\right]_{cl}\subset\int_{\sigma_0}F(d\sigma)+V$$

for every $V\in\mathcal{V}$. Since $\int_{\sigma_0}F(d\sigma)$ is closed, $\int_{\sigma_0}G(d\sigma)\subseteq\int_{\sigma_0}F(d\sigma)$.

5.12. COROLLARY. *If $F(\sigma)$ is $\mathfrak{S}\mathcal{V}$-integrable on $\sigma_0$ and if, for $\sigma\subseteq\sigma_0$, $G(\sigma)\subseteq F(\sigma)_{cl}$, then $G(\sigma)$ is $\mathfrak{S}\mathcal{V}$-integrable on $\sigma_0$ and $\int_{\sigma_0}G(d\sigma)=\int_{\sigma_0}F(d\sigma)$.*

6. **Differential equivalence.** The results of this section parallel similar results obtained by Kolmogoroff for the case of $\mathfrak{X}$ the real numbers. The following definition of differential equivalence is a direct generalization of the Kolmogoroff definition [7, p. 666].

6.1. DEFINITION. *The functions $F(\sigma)$, $G(\sigma)$ are said to be differentially equivalent (we write d.e.) on the set $\sigma_0$ provided for every $V\in\mathcal{V}$, there exists a*

*$\Delta_V$ such that, if $\Delta \geqq \Delta_V$, then $J(F, \sigma_0, \Delta)$, $J(G, \sigma_0, \Delta)$ are summably equal within* $V$. (See Definition 2.1 above.)

6.2. THEOREM. *If $F(\sigma)$, $G(\sigma)$ are d.e. on $\sigma_0$, then the $\mathcal{V}$-integrability of either function on $\sigma_0$ implies the $\mathcal{V}$-integrability of the other and to the same value.*

Suppose $F(\sigma)$ $\mathcal{V}$-integrable on $\sigma_0$. Evidently, for every $V \in \mathcal{V}$, there exists $\Delta_V$ such that $\Delta \geqq \Delta_V$ implies that $J(F, \sigma_0, \Delta)$ is u.s. to $\int_{\sigma_0} F(d\sigma)$ with respect to $V$ and that $J(F, \sigma_0, \Delta)$, $J(G, \sigma_0, \Delta)$ are summably equal within $V$. From this it follows that $J(G, \sigma_0, \Delta)$ is u.s. to $\int_{\sigma_0} F(d\sigma)$ with respect to $2V$ for $\Delta \geqq \Delta_V$. Since $V$ is arbitrary $G(\sigma)$ is $\mathcal{V}$-integrable to $\int_{\sigma_0} F(d\sigma)$ by definition.

6.3. THEOREM. *If $F(\sigma)$, $G(\sigma)$ are $\mathcal{V}$-integrable on $\sigma_0$ to the same value, then they are d.e. on $\sigma_0$.*

6.4. COROLLARY. *If $F(\sigma)$ is $\mathcal{V}$-integrable on every $\sigma \subseteq \sigma_0$, then $F(\sigma)$ and $\int_\sigma F(d\sigma)$ are d.e. on $\sigma_0$.*

In view of the preceding results, one can characterize the (indefinite) $\mathcal{V}$-integral in terms of differential equivalence.

6.5. THEOREM. *In order that a function $I(\sigma)$ be the (indefinite) $\mathcal{V}$-integral of a given function $F(\sigma)$, it is both necessary and sufficient that it be closed (that is, $I(\sigma) = I(\sigma)_{cl}$), completely additive in the sense of Theorem 5.5 and d.e. to $F(\sigma)$ on each $\sigma$.*

6.6. COROLLARY. *In order that a single-valued function $I(\sigma)$ be the (indefinite) $\mathfrak{S}$ $\mathcal{V}$-integral of a given function $F(\sigma)$, it is both necessary and sufficient that it be completely additive in the ordinary sense and d.e. to $F(\sigma)$ on each $\sigma$.*

7. **Transformation of an integrable function.** We now introduce a general type of linear transformation $T(X)$ defined on subsets of $\mathfrak{X}$ and whose values are sets in a similar space $\mathfrak{Y}$. The topology on $\mathfrak{Y}$ will be given by the system of sets $U$ individual elements of which will be denoted by $U$. $T(X)$ will be subject to the following three conditions:

(1) $X_1 \subset X_2$ implies $T(X_1) \subset T(X_2)$.

(2) $T(X)$ is linear, that is, $T(\alpha_1 X_1 + \alpha_2 X_2) = \alpha_1 T(X_1) + \alpha_2 T(X_2)$.

(3) $T(X)$ is continuous in the sense that $U \in U$ implies the existence of a $V_U \in \mathcal{V}$ such that $T(V_U) \subset U$.

The class of transformations described above contains as a special case the ordinary linear continuous point transformations $T(x)$ on $\mathfrak{X}$ to $\mathfrak{Y}$, where $T(X) = \{T(x) \mid x \in X\}$. It also contains the operation of forming the convex $C_oX$ of a set $X$ and the operation of forming the "generalized convex" $C^*(X)$ of a set using the bounded generalized convex operators of G. B. Price [13, p. 7]. Observe that in these last two instances $\mathfrak{X} = \mathfrak{Y}$ and the transformations have the additional property of leaving individual points invariant.

7.1. THEOREM. *If $F(\sigma)$ is $\mathcal{V}$-integrable on $\sigma_0$, then the function $T(F(\sigma))$ is $\mathcal{V}$-integrable on $\sigma_0$ and $\int_{\sigma_0} T(F(d\sigma)) = T[\int_{\sigma_0} F(d\sigma)]_{cl}$.*

Since $T(X)$ is continuous, for arbitrary $U \in \mathcal{V}$ there exists $V_U \in \mathcal{V}$ such that $T(V_U) \subset U$. Also, since $F(\sigma)$ is $\mathcal{V}$-integrable on $\sigma_0$, there exists $\Delta_U$ such that, if $\{\sigma_i\} = \Delta \geqq \Delta_U$, then there exists $\pi_\Delta$ for which $\pi \geqq \pi_\Delta$ implies

$$\sum_\pi F(\sigma_0 \cap \sigma_i) \subset \int_{\sigma_0} F(d\sigma) + V_U, \qquad \int_{\sigma_0} F(d\sigma) \subset \sum_\pi F(\sigma_0 \cap \sigma_i) + V_U.$$

Application of $T$ to these relations and use of (1), (2) give

$$\sum_\pi T(F(\sigma_0 \cap \sigma_i)) \subset T\left[\int_{\sigma_0} F(d\sigma)\right] + U,$$

$$T\left[\int_{\sigma_0} F(d\sigma)\right] \subset \sum_\pi T(F(\sigma_0 \cap \sigma_i)) + U.$$

Therefore $J(T(F), \sigma_0, \Delta)$ is u.s. to $T[\int_{\sigma_0} F(d\sigma)]$ with respect to $U$, which completes the proof.

7.2. COROLLARY. *If $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$, then*(18) *$T(F(\sigma))$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$ and $\int_{\sigma_0} F(T(d\sigma)) = T[\int_{\sigma_0} F(d\sigma)]$.*

7.3. COROLLARY. *If $\mathfrak{X} = \mathfrak{Y}$ and single elements are invariant under $T$, then $\mathfrak{S}\,\mathcal{V}$-integrability of $F(\sigma)$ implies that of $T(F(\sigma))$ and to the same value; that is, $\int_{\sigma_0} T(F(d\sigma)) = \int_{\sigma_0} F(d\sigma)$.*

8. **The $\mathfrak{S}\,\mathcal{V}$-integral in a complete space.** It will be recalled that the definition of the $\mathcal{V}$- and $\mathfrak{S}\,\mathcal{V}$-integrals (Definition 4.1) involves an assumption concerning the existence of the value of the integral in the space. This is, in part, necessitated by a lack of completeness in the space $\mathfrak{X}$. It is the purpose of this section to show that the existence assumption can be dropped in the case of the $\mathfrak{S}\,\mathcal{V}$-integral provided the space $\mathfrak{X}$ is complete relative to $\mathcal{V}$ (see §1 above).

8.1. DEFINITION. *The function $F(\sigma)$ is said to be conditionally $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$ if, for every $V \in \mathcal{V}$, there exists a $\Delta_{V\sigma_0}$ so that $\{\sigma_i^j\} = \Delta^j \geqq \Delta_{V\sigma_0}$ implies the existence of independent $\pi(\Delta^j)$ $(j = 1, 2)$ for which it is true that $\sum_{\pi_1} F(\sigma_0 \cap \sigma_i^1) - \sum_{\pi_2} F(\sigma_0 \cap \sigma_i^2) \subset V$ whenever $\pi_j \geqq \pi(\Delta^j)$ $(j = 1, 2)$.*

8.2. THEOREM. *If $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$, then $F(\sigma)$ is conditionally $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$.*

8.3. THEOREM. *If $\mathfrak{X}$ is complete relative to $\mathcal{V}$ and $F(\sigma)$ is conditionally $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$, then $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$.*

---

(18) Observe that the continuity of $T(X)$ implies that single elements are carried into single elements; that is, $T(X)$ induces a linear continuous point transformation on $\mathfrak{X}$ to $\mathfrak{Y}$.

Let $\Delta_{V\sigma_0}=\{\sigma_{iV}\}$ be the subdivision and $\pi(\Delta_{V\sigma_0})=\pi_V$ the associated set of positive integers given by Definition 8.1. Denote by $x_V$ a particular one of the elements in the set $\sum_{\pi_V}' F(\sigma_0\cap\sigma_{iV})$; then, by Definition 8.1, $J(F, \sigma_0, \Delta)$ is u.s. to $x_V$ with respect to $V$ for every $\Delta\geqq\Delta_{V\sigma_0}$. We now prove that $\{x_V\}$ is a fundamental $\mathcal{V}$-directed set.

Let $V\in\mathcal{V}$ be arbitrary and consider any pair of elements $V_1$, $V_2\in\mathcal{V}$ such that $V_i\subset V/2$ $(i=1, 2)$. For $\Delta\geqq\Delta_{V_1\sigma_0}\Delta_{V_2\sigma_0}$ we have that $J(F, \sigma_0, \Delta)$ is u.s. to $x_{V_1}$ with respect to $V_1$ and to $x_{V_2}$ with respect to $V_2$. It follows directly from this result that $x_{V_1}-x_{V_2}\in V_1+V_2\subset V$ and, hence, that $\{x_V\}$ is a fundamental $\mathcal{V}$-directed set. Let $x_0$ be the limit of this set. It remains to show that $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$ to the value $x_0$.

For arbitrary $V\in\mathcal{V}$ first chose $V_0\subset V/2$ such that $V'\subset V_0$ implies $\pm\{x_{V'}-x_0\}\in V/2$ and then choose $\Delta_{V_0\sigma_0}$ according to Definition 8.1. Then, if $\Delta\geqq\Delta_{V_0\sigma_0}$, $J(F, \sigma_0, \Delta)$ is u.s. to $x_{V_0}$ with respect to $V_0$. But $\pm\{x_{V_0}-x_0\}\in V/2$; therefore $J(F, \sigma_0, \Delta)$ is u.s. to $x_0$ with respect to $V$; that is, $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on $\sigma_0$ to the value $x_0$.

The proof of the following lemma, though not difficult, is somewhat long; so will be omitted.

8.4. LEMMA. *Conditional $\mathfrak{S}\,\mathcal{V}$-integrability on $M$ implies conditional $\mathfrak{S}\,\mathcal{V}$-integrability on every $\sigma$.*

Combining Lemma 8.4 with Theorem 8.3 we have

8.5. THEOREM[19]. *If $\mathfrak{X}$ is complete relative to $\mathcal{V}$ and $F(\sigma)$ is conditionally $\mathfrak{S}\,\mathcal{V}$-integrable on $M$, then $F(\sigma)$ is $\mathfrak{S}\,\mathcal{V}$-integrable on every $\sigma$.*

9. **A convergence theorem for the $\mathfrak{S}\,\mathcal{V}$-integral.** We consider only the $\mathfrak{S}\,\mathcal{V}$-integral in this section and restrict attention to integrable functions $F(\sigma)$ for which $\int_\sigma F(d\sigma)$ is absolutely continuous relative to a given, positive, completely additive measure function $m(\sigma)$. In view of Theorem 3.1, we could replace the above restriction by the stronger condition that $m(\sigma)=0$ imply $F(\sigma)=\theta$.

The following definition gives a generalization of the notion of approximate convergence[20] to functions $F(\sigma)$ of the type being considered here. It is also a generalization of a much stronger type of convergence used by Kolmogoroff [7, p. 665].

9.1. DEFINITION. *A sequence of functions $\{F_n(\sigma)\}$ is said to converge approximately to $F(\sigma)$ relative to $m(\sigma)$ provided, for every integer $n$ and $V\in\mathcal{V}$, there exists a $\sigma(n, V)\in\mathfrak{M}$ and a subdivision $\Delta_{nV}$ such that, for each $V$,*

---

(19) Compare with Phillips' Theorem 4.1 [12, p. 122]. Observe that "completeness with respect to $D$" used by Phillips implies completeness relative to $\mathcal{V}$ (they are, in fact, equivalent; see Footnote 7).

(20) See Definition 12.3 and Theorem 12.4 below.

$\lim_{n\to\infty} m(\sigma(n, V))=0$ *and, for* $\Delta \geqq \Delta_{nV}$, *it is true that* $J(F_n, \sigma, \Delta)$, $J(F, \sigma, \Delta)$ *are summably equal within* $V$ *for every* $\sigma \subseteq M \cap C\sigma(n, V)$.

Observe that, if $\Delta = \{\sigma_i\}$ and $J(F_n, \sigma, \Delta)$, $J(F, \sigma, \Delta)$ are summably equal within $V$ for arbitrary $\sigma \subseteq M \cap C\sigma(n, V)$, then $J(F_n, \sigma \cap \bigcup_\pi \sigma_i, \Delta)$, $J(F, \sigma \cap \bigcup_\pi \sigma_i, \Delta)$ are summably equal within $V$ for *arbitrary* $\pi$ and $\sigma \subseteq M \cap C\sigma(n, V)$. It follows immediately that $\sum_\pi F_n(\sigma \cap \sigma_i)$, $\sum_\pi F(\sigma \cap \sigma_i)$ are equal within $V$ for *arbitrary* $\pi$ and $\sigma \subseteq M \cap C\sigma(n, V)$. We thus obtain a result which is somewhat stronger than summable equality.

The proof of the next theorem will be omitted, since it is essentially contained in the first part of the proof of Theorem 9.5 below.

9.2. THEOREM. *Let* $F_n(\sigma)$ *be* $\mathfrak{S}\mathcal{V}$*-integrable on every* $\sigma$ *and let* $\int_\sigma F_n(d\sigma)$ *be absolutely continuous relative to* $m(\sigma)$ $(n=0, 1, 2, \cdots)$. *Then, if* $F_n(\sigma)$ *converges approximately to* $F_0(\sigma)$ *relative to* $m(\sigma)$, *the following are equivalent*:

(i) $\lim_{n\to\infty} \int_\sigma F_n(d\sigma) = \int_\sigma F_0(d\sigma)$ *uniformly in* $\sigma$.

(ii) $\int_\sigma F_n(d\sigma)$ *are equi-absolutely continuous relative to* $m(\sigma)$.

9.3. DEFINITION. *The function* $F(\sigma)$ *is said to be* $\mathfrak{S}\mathcal{V}$*-integrable uniformly in* $\sigma$ *provided* $F(\sigma)$ *is* $\mathfrak{S}\mathcal{V}$*-integrable on every* $\sigma$ *and, for each* $V \in \mathcal{V}$, *there exists* $\Delta_V$ *independent of* $\sigma$ *such that, if* $\Delta \geqq \Delta_V$, *then* $J(F, \sigma, \Delta)$ *is u.s. to* $\int_\sigma F(d\sigma)$ *with respect to* $V$ *uniformly in* $\sigma$.

9.4. LEMMA. *Let* $F(\sigma)$ *be* $\mathfrak{S}\mathcal{V}$*-integrable uniformly in* $\sigma$ *and let* $\sigma_0$ *be such that* $\pm\int_{\sigma\cap\sigma_0} F(d\sigma) \in V$ *for all* $\sigma$. *Then there exists* $\Delta_V$ *such that* $\{\sigma_i\} = \Delta \geqq \Delta_V$ *implies* $\pm\sum_{\pi'} F(\sigma_0 \cap \sigma_i) \subset 2V$ *for arbitrary* $\pi'$.

From the definition of uniform $\mathfrak{S}\mathcal{V}$-integrability, there exists $\Delta_V$ such that $\{\sigma_i\} = \Delta \geqq \Delta_V$ implies that $J(F, \sigma, \Delta)$ is u.s. to $\int_\sigma F(d\sigma)$ with respect to $V$ uniformly in $\sigma$, that is, there exists $\pi_\Delta$ independent of $\sigma$ such that, if $\pi \geqq \pi_\Delta$,

$$\pm\left\{ \sum_\pi F(\sigma \cap \sigma_i) - \int_\sigma F(d\sigma)\right\} \subset V. \tag{1}$$

Now let $\pi'$ be arbitrary and set $\sigma = \sigma_0 \cap (\bigcup_{\pi'} \sigma_i)$, $\pi = \pi' \cup \pi_\Delta$ in (1). Since $\pm\int_{\sigma\cap\sigma_0} F(d\sigma) \in V$ for all $\sigma$, this completes the proof.

9.5. THEOREM. *Let* $\mathfrak{X}$ *be sequentially complete*, $F_n(\sigma)$ $\mathfrak{S}\mathcal{V}$*-integrable uniformly in* $\sigma$, *and* $\int_\sigma F_n(d\sigma)$ *absolutely continuous relative to* $m(\sigma)$ $(n=1, 2, \cdots)$. *Then, if* $\{F_n(\sigma)\}$ *converges approximately to* $F(\sigma)$ *relative to* $m(\sigma)$, *the following are equivalent*:

(i) $F(\sigma)$ *is* $\mathfrak{S}\mathcal{V}$*-integrable uniformly in* $\sigma$ *and* $\lim_{n\to\infty} \int_\sigma F_n(d\sigma) = \int_\sigma F(d\sigma)$ *uniformly in* $\sigma$.

(ii) $\lim_{n\to\infty} \int_\sigma F_n(d\sigma)$ *exists for every* $\sigma$.

(iii) $\int_\sigma F_n(d\sigma)$ *are equi-absolutely continuous relative to* $m(\sigma)$.

That (i) implies (ii) is trivial and (ii) implies (iii) by Theorem 3.2. We prove that (iii) implies (i).

Let $V \in \mathcal{V}$ be arbitrary and set $\sigma_{mn} = \sigma(m, V) \cup \sigma(n, V)$, where $\sigma(m, V)$, $\sigma(n, V)$ are given by Definition 9.1. If $\Delta \geqq \Delta_{mV} \cdot \Delta_{nV}$, where $\Delta_{mV}$, $\Delta_{nV}$ are given by Definition 9.1, and if $\sigma \subseteq M \cap C\sigma_{mn}$, then $J(F_m, \sigma, \Delta)$, $J(F_n, \sigma, \Delta)$ are each summably equal to $J(F, \sigma, \Delta)$ within $V$. It follows that $J(F_m, \sigma, \Delta)$, $J(F_n, \sigma, \Delta)$ are summably equal within $2V$. Now an application of Corollary 5.9 gives

$$\int_{\sigma \cap C\sigma_{mn}} F_m(d\sigma) - \int_{\sigma \cap C\sigma_{mn}} F_n(d\sigma) \in (2V)_{cl} \subset 3V.$$

This holds for arbitrary $\sigma$ and all $m, n$. Using (iii) ,we obtain $n_V$ such that $m, n \geqq n_V$ implies $\pm \int_{\sigma \cap \sigma_{mn}} F_k(d\sigma) \in V$ for arbitrary $\sigma$ and all $k$. Therefore, if $m, n \geqq n_V$,

$$\int_\sigma F_m(d\sigma) - \int_\sigma F_n(d\sigma) \in 3V + 2V \subset 6V,$$

where $\sigma$ is arbitrary and $n_V$ obviously does not depend on $\sigma$. It follows that $\{\int_\sigma F_n(d\sigma)\}$ is a fundamental sequence uniformly in $\sigma$. Since $\mathfrak{X}$ is sequentially complete, there exists $I(\sigma) \in \mathfrak{X}$ such that $\lim_{n\to\infty} \int_\sigma F_n(d\sigma) = I(\sigma)$ uniformly in $\sigma$. It remains to show that $F(\sigma)$ is $\mathfrak{S}\mathcal{V}$-integrable uniformly in $\sigma$ to the value $I(\sigma)$.

Let $V \in \mathcal{V}$ be arbitrary and select a subsequence of the $F_n(\sigma)$, which we continue to denote by $\{F_n(\sigma)\}$, having the following two properties:

(a) $\pm\{\int_\sigma F_1(d\sigma) - I(\sigma)\} \in 2^{-3} V$ for all $\sigma$.

(b) There exist $\tau_n \in \mathfrak{M}$ such that([21]) $m(\tau_n) < \delta(2^{-n-4}V)$, $\tau_n \supset \tau_{n+1}$, and also there exist $\Delta_{nV}$ such that $\Delta \geqq \Delta_{nV}$ implies $J(F_n, \sigma, \Delta)$, $J(F, \sigma, \Delta)$ are summably equal within $\pm 2^{-n-2}V$ for all $\sigma \subseteq M \cap C\tau_n$.

Set $\sigma_1^0 = M \cap C\tau_1$ and $\sigma_n^0 = \tau_{n-1} \cap C\tau_n$ for $n \geqq 2$, and consider the subdivision $\Delta^0 = \{\sigma_i^0\}$. Since $m(\sigma \cap \sigma_n^0) < \delta(2^{-n-3}V)$ (for $n \geqq 2$), it follows that $\pm \int_{\sigma \cap \sigma_n^0} F_n(d\sigma) \in 2^{-n-3}V$ for arbitrary $\sigma$. Hence, by Lemma 9.4, there exists $\Delta^n \geqq \Delta^0 \Delta_{nV}$ such that $\{\sigma_i\} = \Delta \geqq \Delta^n$ implies

$$\pm \sum_{\pi'} F_n(\sigma \cap \sigma_n^0 \cap \sigma_i) \subset V/2^{n+2},$$

for arbitrary $\pi'$, $\sigma$ and $n \geqq 2$. Out of property (b) and the remark following Definition 9.1, it follows that $\sum_{\pi'} F_n(\sigma \cap \sigma_n^0 \cap \sigma_i)$, $\sum_{\pi'} F(\sigma \cap \sigma_n^0 \cap \sigma_i)$ are equal within $\pm 2^{-n-2}V$ and, hence, that

$$\pm \sum_{\pi'} F(\sigma \cap \sigma_n^0 \cap \sigma_i) \subset V/2^{n+1}.$$

This result holds for arbitrary $n \geqq 2$, $\sigma$, $\pi'$ and $\{\sigma_i\} = \Delta \geqq \Delta^n$.

Now define $\Delta^1$ such that $\Delta \geqq \Delta^1$ implies $J(F, \sigma, \Delta)$ u.s. to $\int_\sigma F_1(d\sigma)$ with respect to $2^{-4}V$ uniformly in $\sigma$, and let $\Delta_V$ be the sum of the subdivisions $\Delta^n$

---

([21]) $\delta(eV) > 0$ is chosen so that, if $m(\sigma) < \delta(eV)$, then $\pm \int_\sigma F_k(d\sigma) \in eV$ for all $k$.

($n=1, 2, \cdots$) over the subdivision $\Delta^0$ (see §4 above). For $\{\sigma_i\}=\Delta \geqq \Delta_V$ and arbitrary $\pi'$, $\sigma$, we have

$$(2) \qquad \pm \sum_{\pi'} F(\sigma \cap \tau_1 \cap \sigma_i) = \pm \sum_{n=2}^{\infty} \sum_{\pi'} F(\sigma \cap \sigma_n^0 \cap \sigma_i) \subset \sum_{n=2}^{N_{\pi'}} V/2^{n+1} \subset V/2,$$

where $N_{\pi'}$ is the largest $n$ for which $\sigma_n^0 \cap (\bigcup_{\pi'} \sigma_i) \neq 0$.

If $\Delta \geqq \Delta_V$, then $J(F_1, \sigma \cap \sigma_1^0, \Delta)$ is u.s. to $\int_{\sigma\cap\sigma^0} F_1(d\sigma)$ with respect to $2^{-4}V$ uniformly in $\sigma$. Moreover, since $m(\tau_1) < \delta(2^{-5}V)$, it follows that $\pm\int_{\sigma\cap\tau_1} F_1(d\sigma) \in 2^{-5}V$ and, hence, that $J(F_1, \sigma \cap \sigma_1^0, \Delta)$ is u.s. to $\int_\sigma F_1(d\sigma)$ with respect to $2^{-3}V$ uniformly in $\sigma$. Applying (a), we obtain $J(F_1, \sigma \cap \sigma_1^0, \Delta)$ u.s. to $I(\sigma)$ with respect to $V/2$ and, applying (b) again, we have $J(F, \sigma \cap \sigma_1^0, \Delta)$ u.s. to $I(\sigma)$ with respect to $V/2$ uniformly in $\sigma$. From this last result and (2) it follows that $J(F, \sigma, \Delta)$ is u.s. to $I(\sigma)$ with respect to $V$ uniformly in $\sigma$, which completes the proof of Theorem 9.5.

## PART III. INTEGRATION WITH RESPECT TO A "BILINEAR" FUNCTION(22)

10. **The $\mathcal{V}_B$-integral.** Let $\mathfrak{X}$, as usual, be a convex linear topological space and let $\mathfrak{Y}$ be simply a linear space. We introduce a "bilinear" function $B[y, \sigma]$ subject to the following four conditions:

B1. *For every $y \in \mathfrak{Y}$ and*(23) *$\sigma \in \mathfrak{M}$, $B[y, \sigma]$ is a unique element of $\mathfrak{X}$.*

B2. *$B[y, \sigma]$ is linear (not necessarily continuous) in $y$ for each $\sigma$; that is, $B[\alpha_1 y_1 + \alpha_2 y_2, \sigma] = \alpha_1 B[y_1, \sigma] + \alpha_2 B[y_2, \sigma]$.*

B3. *For each $y$, $B[y, \sigma]$ is a completely additive function of $\sigma$.*

B4. *There exists a real number $\beta \geqq 1$ such that*(24)

$$\sum_{i=1}^{m} B[Y_i, \sigma_i] \subset V \quad \textit{implies} \quad \sum_{i=1}^{m} \sum_{j=1}^{n_i} B[Y_i, \sigma_i^j] \subset \beta V,$$

*where $Y_i \subset \mathfrak{Y}$, $\sigma_i \cap \sigma_j = 0$ $(i \neq j)$, $\sigma_i = \bigcup_{j=1}^{n_i} \sigma_i^j$, $\sigma_i^j \cap \sigma_i^k = 0$ $(j \neq k)$.*

The functions $y(\sigma)$ to be considered in this part will be multi-valued and defined on $\mathfrak{M}$ to $\mathfrak{Y}$. They will be subject without exception to the restriction that $\sigma_1 \subseteq \sigma_2$ shall imply $y(\sigma_1) \subseteq y(\sigma_2)$. Such functions are described as *contractive*(25). If $\Delta = \{\sigma_i\}$ is an arbitrary subdivision, the sequence of sets $\{B[y(\sigma \cap \sigma_i), \sigma \cap \sigma_i]\}$ will be denoted by the symbol $J_B(y, \sigma, \Delta)$.

10.1. DEFINITION. *$y(\sigma)$ is said to be $\mathcal{V}_B$-integrable on $\sigma_0$ provided there exists an element $I_B(y, \sigma_0) \in \mathfrak{X}$ such that, for every $V \in \mathcal{V}$, there exists a $\Delta_{\sigma_0 V}$ for which $J_B(y, \sigma_0, \Delta_{\sigma_0 V})$ is u.s. to $I_B(y, \sigma_0)$ with respect to $V$. $I_B(y, \sigma_0)$ is the value of the integral and we write $I_B(y, \sigma_0) = \int_{\sigma_0} B[y, d\sigma]$.*

(22) The general idea of considering integration with respect to a "bilinear" function was suggested by T. H. Hildebrandt. This paper represents a development from that idea.

(23) See Footnote 14 above.

(24) If $Y \subset \mathfrak{Y}$, then $B[Y, \sigma] = \{B[y, \sigma] \mid y \in Y\}$.

(25) Observe that, if $y(t)$ is a point function, the associated set function $y(\sigma) = \{y(t) \mid t \in \sigma\}$ is contractive.

Observe that this definition is weaker in form than Definition 4.1, because the assumption here is that $J_B(y, \sigma_0, \Delta)$ be u.s. to $I_B(y, \sigma_0)$ only for $\Delta = \Delta_{\sigma_0 V}$ rather than $\Delta \geqq \Delta_{\sigma_0 V}$. The weakening of the definition of integrability is balanced by the conditions on $B[y, \sigma]$. Definition 10.1 is essentially that used by R. S. Phillips [12, p. 118] and the integral obtained here will be seen to reduce to his as a special case (Theorem 15.3).

10.2. LEMMA. *Let* $Y_i \subset \mathfrak{Y}$, $\sigma_i \cap \sigma_j = 0$ $(i \neq j)$, $\sigma_i = \bigcup_{j=1}^{\infty} \sigma_i^j$, $\sigma_i^j \cap \sigma_i^k = 0$ $(j \neq k)$; *then*

$$\pm \left\{ x + \sum_{i=1}^{m} B[Y_i, \sigma_i] \right\} \subset V$$

*implies the existence of a* $\pi_0$ *independent of* $x$ *such that, if* $\pi_i \geqq \pi_0$,

$$\pm \left\{ x + \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] \right\} \subset 4\beta V.$$

We have immediately that

$$\sum_{i=1}^{m} B[Y_i - Y_i, \sigma_i] \subset 2V;$$

hence, by condition B4,

$$\sum_{i=1}^{m} \left\{ \sum_{j \in \pi_i} B[Y_i - Y_i, \sigma_i^j] + B\left[ Y_i - Y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right] \right\} \subset 2\beta V,$$

where the $\pi_i$ are completely arbitrary. This can be written in the form

$$\begin{aligned} \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] + \sum_{i=1}^{m} B\left[ Y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right] - \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] \\ - \sum_{i=1}^{m} B\left[ Y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right] \subset 2\beta V. \end{aligned} \tag{1}$$

This gives

$$\pm \left\{ \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] + \sum_{i=1}^{m} B\left[ Y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right] - \sum_{i=1}^{m} B[Y_i, \sigma_i] \right\} \subset 2\beta V.$$

It follows by the hypothesis of the lemma that

$$\pm \left\{ x + \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] + \sum_{i=1}^{m} B\left[ Y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right] \right\} \subset 2\beta V + V. \tag{2}$$

Now let $y_i$ be some particular element of $Y_i$; then from condition B3 it is evident that there exists a $\pi_0$ such that $\pi_i \geqq \pi_0$ implies

$$\pm \sum_{i=1}^{m} B\left\{ y_i, \bigcup_{j \notin \pi_i} \sigma_i^j \right\} \in V. \tag{3}$$

Relations (2) and (3) together give

$$\pm \left\{ x + \sum_{i=1}^{m} \sum_{j \in \pi_i} B[Y_i, \sigma_i^j] \right\} \subset 2\beta V + V + V \subset 4\beta V,$$

which completes the proof.

10.3. Lemma. *If $J_B(y, \sigma, \Delta_0)$ is u.s. to $I_B(y, \sigma)$ with respect to $V$, then $J_B(y, \sigma, \Delta)$ is u.s. to $I_B(y, \sigma)$ with respect to $7\beta V$ for all $\Delta \geqq \Delta_0$.*

There is no loss in taking $\sigma = M$. Let $\Delta_0 = \{\sigma_i\}$; then by hypothesis there exists $\pi_0$ such that $\pi \geqq \pi_0$ gives

(1) $$\pm \left\{ \sum_{\pi} B[y(\sigma_i), \sigma_i] - I_B(y, M) \right\} \subset V.$$

Consider $\{\sigma_i^j\} = \Delta \geqq \Delta_0$, where $\sigma_i = \bigcup_{j=1}^{\infty} \sigma_i^j$. By Lemma 10.2 there exists $\pi_0'$ such that $\pi_i \geqq \pi_0'$ implies

(2) $$\pm \left\{ \sum_{i \in \pi_0} \sum_{j \in \pi_i} B[y(\sigma_i^j), \sigma_i^j] - I_B(y, M) \right\} \subset 4\beta V.$$

Now let $\nu_0$ denote those integer pairs $(i, j)$ for which $i \in \pi_0$ and $j \in \pi_0'$, and consider any finite set of integer pairs $\nu$ which contains $\nu_0$; that is, $\nu \geqq \nu_0$. Set $\nu' = \{(i, j) \mid (i, j) \in \nu, i \in \pi_0\}$ and $\nu'' = \{(i, j) \mid (i, j) \in \nu, i \notin \pi_0\}$. It follows from (2) that

(3) $$\pm \left\{ \sum_{\nu'} B[y(\sigma_i^j), \sigma_i^j] - I_B(y, M) \right\} \subset 4\beta V.$$

Moreover, from (1) we have

(4) $$\pm \sum_{\pi'} B[y(\sigma_i), \sigma_i] \subset 2V,$$

for arbitrary $\pi'$ such that $\pi' \cap \pi_0 = 0$. Because of the arbitrary character of $\pi'$ in (4), we can write

(5) $$\pm \sum_{\pi'} B[y(\sigma_i) \cup \theta, \sigma_i] \subset 2V.$$

Now let $\pi' = \{i \mid (i, j) \in \nu'' \text{ for some } j\}$, $\pi_i' = \{j \mid (i, j) \in \nu''\}$ and apply B4 to (5) to obtain

$$\pm \left\{ \sum_{\nu''} B[y(\sigma_i) \cup \theta, \sigma_i^j] + \sum_{\pi'} B\left[ y(\sigma_i) \cup \theta, \bigcup_{j \notin \pi_i'} \sigma_i^j \right] \right\} \subset 2\beta V.$$

From this it follows that

(6) $$\pm \sum_{\nu''} B[y(\sigma_i^j), \sigma_i^j] \subset 2\beta V.$$

Combining (3) and (6) we obtain

$$\pm\left\{\sum_{p} B[y(\sigma_i^j), \sigma_i^j] - I_B(y, M)\right\} \subset 4\beta V + 2\beta V \subset 7\beta V.$$

As a consequence of Lemma 10.3 we have

10.4. THEOREM. *$\mathcal{V}_B$-integrability of $y(\sigma)$ is equivalent to* $\mathfrak{S}\,\mathcal{V}$*-integrability of* $F(\sigma)=B[y(\sigma), \sigma]$.

Investigation of the form in which the basic properties of the $\mathfrak{S}\mathcal{V}$-integral appear in the special case of the $\mathcal{V}_B$-integral will be left to the reader.

The following lemma is in preparation for the proof that $\mathcal{V}_B$-integrability of $y(\sigma)$ for every $\sigma$ implies $\mathfrak{S}\,\mathcal{V}$-integrability of $F(\sigma)$ uniformly in $\sigma$.

10.5. LEMMA. *If for a given $\Delta=\{\sigma_i\}$ there exists $\pi_\Delta$ such that $\pi_i \geqq \pi_\Delta$ $(i=1, 2)$ implies*

$$\sum_{\pi_1} B[y(\sigma_i), \sigma_i] - \sum_{\pi_2} B[y(\sigma_i), \sigma_i] \subset V, \tag{1}$$

*then for arbitrary $\sigma$ and $\pi_i \geqq \pi_\Delta$ $(i=1, 2)$*

$$\sum_{\pi_1} B[y(\sigma \cap \sigma_i), \sigma \cap \sigma_i] - \sum_{\pi_2} B[y(\sigma \cap \sigma_i), \sigma \cap \sigma_i] \subset 4\beta V.$$

Taking $\pi_1=\pi_2=\pi_\Delta$, we obtain from (1)

$$\sum_{\pi_\Delta} B[y(\sigma_i) - y(\sigma_i), \sigma_i] \subset V.$$

An application of B4 gives

$$\sum_{\pi_\Delta} \{B[y(\sigma_i) - y(\sigma_i), \sigma \cap \sigma_i] - B[y(\sigma_i) - y(\sigma_i), \sigma_i \cap C\sigma]\} \subset \beta V.$$

Since $\theta \in y(\sigma_i) - y(\sigma_i)$, it follows that

$$\sum_{\pi_\Delta} B[y(\sigma \cap \sigma_i), \sigma \cap \sigma_i] - \sum_{\pi_\Delta} B[y\ (\sigma \cap \sigma_i), \sigma \cap \sigma_i] \subset \beta V. \tag{2}$$

Again from (1) we have for arbitrary $\pi' \cap \pi_\Delta = 0$

$$\pm \sum_{\pi'} B[y(\sigma_i), \sigma_i] \subset V,$$

which, because of the arbitrary character of $\pi'$, implies (as in the proof of Lemma 10.3)

$$\pm \sum B[y(\sigma \cap \sigma_i), \sigma \cap \sigma_i] \subset \beta V.$$

Combining this last result with (2) completes the proof of the lemma.

10.6. THEOREM. *$\mathcal{V}_B$-integrability of $y(\sigma)$ on every $\sigma$ is equivalent to* $\mathfrak{S}\mathcal{V}$*-integrability of $F(\sigma)=B[y(\sigma), \sigma]$ uniformly in $\sigma$.*

Because of Theorem 10.4, for every $V\in\mathcal{V}$ there exists $\Delta_V$ such that, if $\{\sigma_i\}=\Delta\geqq\Delta_V$, then there exists $\pi_\Delta$ for which $\pi\geqq\pi_\Delta$ implies

$$\pm\left\{\sum_\pi B[y(\sigma_i),\sigma_i]-\int_M B[y,d\sigma]\right\}\subset V/2.$$

Therefore, if $\pi_i\geqq\pi_\Delta$ $(i=1,2)$,

$$\sum_{\pi_1} B[y(\sigma_i),\sigma_i]-\sum_{\pi_2} B[y(\sigma_i),\sigma_i]\subset V.$$

From Lemma 10.5 it follows that

$$(1)\qquad \pm\left\{\sum_{\pi_1} B[y(\sigma\cap\sigma_i),\sigma\cap\sigma_i]-\sum_{\pi_2} B[y(\sigma\cap\sigma_i),\sigma\cap\sigma_i]\right\}\subset 4\beta V$$

for arbitrary $\sigma$ and $\pi_i\geqq\pi_\Delta$ $(i=1,2)$. Now, for a particular $\sigma$ choose $\Delta_{\sigma V}$ such that $\Delta'\geqq\Delta_{\sigma V}$ implies $J_B(y,\sigma,\Delta')$ u.s. to $\int_\sigma B[y,d\sigma]$ with respect to $V$. Let $\{\sigma_i'\}=\Delta'=\Delta\Delta_{\sigma V}$, then it follows that there exists $\pi_{\Delta'}$ such that, if $\pi'\geqq\pi_{\Delta'}$,

$$(2)\qquad \pm\left\{\sum_{\pi'} B[y(\sigma\cap\sigma_i'),\sigma\cap\sigma_i']-\int_\sigma B[y,d\sigma]\right\}\subset V.$$

Now in (1) choose $\pi_2\geqq\pi_\Delta$ such that $\mathsf{U}_{\pi_{\Delta'}}\sigma_i'\subset\mathsf{U}_{\pi_2}\sigma_i$. Then, since $\Delta'\geqq\Delta$, we can apply Lemma 10.2 to (1) and obtain the existence of a $\pi_2'\geqq\pi_{\Delta'}$ such that

$$\pm\left\{\sum_{\pi_1} B[y(\sigma\cap\sigma_i),\sigma\cap\sigma_i]-\sum_{\pi_2'} B[y(\sigma\cap\sigma_i'),\sigma\cap\sigma_i']\right\}\subset 16\beta^2V.$$

This result with (2) yields

$$\pm\left\{\sum_{\pi} B[y(\sigma\cap\sigma_i),\sigma\cap\sigma_i]-\int_\sigma B[y,d\sigma]\right\}\subset 18\beta^2V,$$

for arbitrary $\pi\geqq\pi_\Delta$. Since $\pi_\Delta$ does not depend on $\sigma$, the proof is complete.

11. **The $\mathcal{V}_B$-integral in a complete space.** It has already been observed that the definition of the $\mathcal{V}_B$-integral is weaker in form than the definition of the $\mathfrak{S}\,\mathcal{V}$-integral. Similarly, conditional $\mathcal{V}_B$-integrability can be defined in a weaker form than conditional $\mathfrak{S}\,\mathcal{V}$-integrability.

11.1. DEFINITION. *$y(\sigma)$ is said to be conditionally $\mathcal{V}_B$-integrable on $\sigma_0$ provided for every $V\in\mathcal{V}$ there exists $\Delta_{\sigma_0V}=\{\sigma_i\}$ and $\pi_{\sigma_0V}$ such that, if $\pi_i\geqq\pi_{\sigma_0V}$ $(i=1,2)$, then*

$$\sum_{\pi_1} B[y(\sigma_0\cap\sigma_i),\sigma_0\cap\sigma_i]-\sum_{\pi_2} B[y(\sigma_0\cap\sigma_i),\sigma_0\cap\sigma_i]\subset V.$$

The following theorem follows easily from Lemma 10.2.

11.2. THEOREM. *Conditional $\mathcal{V}_B$-integrability of $y(\sigma)$ is equivalent to conditional $\mathfrak{S}\,\mathcal{V}$-integrability of $F(\sigma)=B[y(\sigma),\sigma]$.*

Out of Theorems 11.2 and 8.5 we obtain

11.3. THEOREM. *If $\mathfrak{X}$ is complete relative to $\mathcal{V}$ and $y(\sigma)$ is conditionally $\mathcal{V}_B$-integrable on $M$, then $y(\sigma)$ is $\mathcal{V}_B$-integrable on every $\sigma$.*

12. **Absolute continuity and a convergence theorem for the $\mathcal{V}_B$-integral.** In this section we assume given a positive completely additive measure function $m(\sigma)$ defined over $\mathfrak{M}$ such that $m(\sigma)=0$ implies $B[y, \sigma]=\theta$ for every $y\in\mathfrak{Y}$. We have immediately that $m(\sigma)=0$ implies $\int_\sigma B[y, d\sigma]=\theta$, where $y(\sigma)$ is arbitrary. This remark plus Theorem 10.4, Corollary 5.6 and Theorem 3.1 enables us to state

12.1. THEOREM. *If $y(\sigma)$ is $\mathcal{V}_B$-integrable on every $\sigma$, then $\int_\sigma B[y, d\sigma]$ is completely additive and absolutely continuous relative to $m(\sigma)$.*

Throughout this and the following section we assume a topology([26]) on the space $\mathfrak{Y}$ given by the system of neighborhoods $\mathcal{U}$ individual elements of which will be denoted by $U$. Also $B[y, \sigma]$ will be subject to the following condition in addition to B1–B4.

B5. *$B[y, M]$ is continuous on $\mathfrak{Y}$ to $\mathfrak{X}$; that is, for every $V\in\mathcal{V}$ there exists $U_V\in \mathcal{U}$ such that $B[U_V, M]\subset V$.*

12.2. THEOREM. *If $B[y, \sigma]$ satisfies* B1–B5, *then $B[y, \sigma]$ is continuous for each $\sigma$ and uniformly in $\sigma$.*

Let $V\in\mathcal{V}$ and choose $U_V$ such that $B[U_V, M]\subset V/\beta$. Applying B4 we get $B[U_V, \sigma]+B[U_V, M\cap C\sigma]\subset V$, where $\sigma$ is arbitrary. Since $\theta\in U_V$, we have $B[U_V, \sigma]\subset V$. But $U_V$ does not depend on $\sigma$; hence $B[y, \sigma]$ is continuous uniformly in $\sigma$.

12.3. DEFINITION. *The sequence of functions $\{y_n(\sigma)\}$ is said to converge approximately*([27]) *to $y(\sigma)$ relative to $m(\sigma)$ provided, for every $n$, $U$, there exists $\sigma(n, U)\in\mathfrak{M}$ such that* $\lim_{n\to\infty} m(\sigma(n, U))=0$ *for each $U$ and for $\sigma\subseteq M\cap C\sigma(n, U)$ it is true that the sets $y_n(\sigma)$, $y(\sigma)$ are equal within $U$.*

12.4. THEOREM. *If $y_n(\sigma)$ converges approximately to $y(\sigma)$ according to Definition* 12.3, *then $F_n(\sigma)=B[y_n(\sigma), \sigma]$ converges approximately to $F(\sigma)=B[y(\sigma), \sigma]$ according to Definition* 9.1.

Given $V\in\mathcal{V}$, because of Theorem 12.2 we can choose $U_V$ such that $B[U_V, \sigma]\subset V/\beta$ for every $\sigma$. It follows immediately from B4 that $\sum_{i=1}^{m} B[U_V, \sigma_i]\subset V$ for arbitrary disjoint $\sigma_i$. Now, if $\sigma_i\subseteq M\cap C\sigma(n, U_V)$, then

$$y_n(\sigma_i) \subset y(\sigma_i) + U_V.$$

Applying $B[y, \sigma_i]$ to this relation and adding, we obtain

([26]) $\mathfrak{Y}$ is not necessarily assumed to be convex.

([27]) This definition is a generalization of the one used by Phillips [12, p. 125].

$$\sum_{i=1}^{m} B[y_n(\sigma_i), \sigma_i] \subset \sum_{i=1}^{m} B[y(\sigma_i), \sigma_i] + V.$$

In a similar manner we obtain

$$\sum_{i=1}^{m} B[y(\sigma_i), \sigma_i] \subset \sum_{i=1}^{m} B[y_n(\sigma_i), \sigma_i] + V.$$

It follows from these relations that $J_B(y_n, \sigma, \Delta)$, $J_B(y, \sigma, \Delta)$ are summably equal within $V$ for *every* $\Delta$ and $\sigma \subseteq M \cap C\sigma(n, U_V)$. Since $\lim_{n\to\infty} m(\sigma(n, U_V)) = 0$, the proof is complete.

Collecting the results of Theorems 10.6, 12.4, 9.5, we can state

12.5. THEOREM. *Let $\mathfrak{X}$ be sequentially complete, $y_n(\sigma)$ be $\mathcal{V}_B$-integrable on each $\sigma$ $(n=1, 2, \cdots)$, and $y_n(\sigma)$ converge approximately to $y(\sigma)$. Then the following are equivalent*:

(i) *$y(\sigma)$ is $\mathcal{V}_B$-integrable on each $\sigma$ and* $\lim_{n\to\infty} \int_\sigma B[y_n, d\sigma] = \int_\sigma B[y, d\sigma]$ *uniformly in $\sigma$.*

(ii) $\lim_{n\to\infty} \int_\sigma B[y_n, d\sigma]$ *exists for each $\sigma$.*

(iii) $\int_\sigma B[y_n, d\sigma]$ *are equi-absolutely continuous relative to $m(\sigma)$.*

13. **Measurable functions. An existence theorem for the $\mathcal{V}_B$-integral.** The following definition of measurability for functions $y(\sigma)$ of the type being considered here is a generalization of a definition given by Price [13, p. 25] for single-valued point functions with values in a Banach space. We are interested here only in generalizing Theorem (16.1) of [13] to the $\mathcal{V}_B$-integral, where $B[y, \sigma]$ is subject to all of the five conditions B1–B5.

13.1. DEFINITION. *The function $y(\sigma)$ is said to be measurable ($\mathfrak{M}$) on the set $\sigma_0$ provided, for every set $Y$ dense in $y(\sigma_0)$, $y \in Y$ and $U \in \mathcal{V}$ implies the existence of $\sigma_{yU} \in \mathfrak{M}$ such that $y(\sigma_{yU}) \subset y + U$ and such that $\sigma_0 = \bigcup_{y\in Y} \sigma_{yU}$.*

The next definition gives a generalization of a familiar condition frequently imposed on Lebesgue measurable functions to insure the existence of a finite Lebesgue integral.

13.2. DEFINITION. *The function $y(\sigma)$ is said to be B-summable provided, for every $V \in \mathcal{V}$, there exists $\Delta_V$ such that, if $\{\sigma_i\} = \Delta \geqq \Delta_V$, then there exists $\pi_\Delta$ for which $\pi \cap \pi_\Delta = 0$ implies $\pm \sum_\pi B[y(\sigma_i), \sigma_i] \subset V$.*

Observe that, if $y(\sigma)$ is $B$-summable, then $J_B(y, M, \Delta)$ is u.s. to $\sum_{\pi_\Delta} B[y(\sigma_i), \sigma_i]$ with respect to $V$ for $\Delta \geqq \Delta_V$. Also it can be proved that $y(\sigma)$ is $B$-summable if there exists $\Delta_V$ with the property that, for each $\Delta \geqq \Delta_V$, there is a *bounded*[28] set $I_\Delta$ such that $J_B(y, M, \Delta)$ is u.s. to $I_\Delta$ with respect to $V$.

[28] A set $X \subset \mathfrak{X}$ is said to be *bounded* provided, for every $V \in \mathcal{V}$, there exists $\alpha > 0$ such that $X \subset \alpha V$ [11].

A function $y(\sigma)$ is said to be *separable* provided the set $y(M)$ is separable. *Almost separable* $(B)$ will mean that there exists a set $\sigma_0$ of $B$-measure zero (that is[29], $B[y, \sigma_0]=\theta$ for every $y\in\mathfrak{Y}$) such that $y(M\cap C\sigma_0)$ is separable [13, p. 25].

13.3. THEOREM. *Let $\mathfrak{X}$ be complete relative to $\mathcal{V}$ and let $y(\sigma)$ be $B$-summable, almost separable $(B)$ and measurable $(\mathfrak{M})$ on the set $M\cap C\sigma_0$, where $\sigma_0$ is the set of $B$-measure zero such that $y(M\cap C\sigma_0)$ is separable. Then $y(\sigma)$ is $\mathcal{V}_B$-integrable on each $\sigma$.*

It may as well be assumed at the outset that $y(\sigma)$ is separable. Also, in view of Theorem 11.3, it will be sufficient to prove $y(\sigma)$ conditionally $\mathcal{V}_B$-integrable on $M$.

Let $V\in\mathcal{V}$ be arbitrary and choose $U_V\in\mathcal{V}$ so that $B[U_V, \sigma]\subset V/\beta$ for all $\sigma$. Then, for arbitrary disjoint $\sigma_i$, it follows by condition B4 that

$$\sum_{i=1}^{m} B[U_V, \sigma_i] \subset V. \tag{1}$$

Since $y(\sigma)$ is measurable $(\mathfrak{M})$ there exists $\sigma_n^0\in\mathfrak{M}$ such that $y(\sigma_n^0)\subset y_n+U'$ and $M=\bigcup_{n=1}^{\infty}\sigma_n^0$, where $\{y_n\}$ is the separating sequence for $y(M)$ and $U'\in\mathcal{V}$ is chosen so that $\pm 2U'\subset U$. Let $\Delta_V$ be the subdivision given by Definition 13.2 and choose $\{\sigma_i\}=\Delta_V'\geqq\Delta_V$ such that each $\sigma_i$ is contained in one of the sets $\sigma_n^0$. Observe that $y(\sigma_i)-y(\sigma_i)\subset U_V$ for every $i$.

Since $\{\sigma_i\}\geqq\Delta_V$, there exists $\pi_V$ such that $\pi'\cap\pi_V=0$ implies $\pm\sum_{\pi'}B[y(\sigma_i), \sigma_i]\subset V$. For arbitrary $\pi_i\geqq\pi_V$ $(i=1, 2)$ set $\pi_i=\pi_V\cup\pi_i'$ where $\pi_i'\cap\pi_V=0$. Then

$$\begin{aligned}\sum_{\pi_1} B[y(\sigma_i), \sigma_i] &- \sum_{\pi_2} B[y(\sigma_i), \sigma_i] \\ &= \sum_{\pi_V} B[y(\sigma_i) - y(\sigma_i), \sigma_i] + \sum_{\pi_1'} B[y(\sigma_i), \sigma_i] - \sum_{\pi_2'} B[y(\sigma_i), \sigma_i] \\ &\subset \sum_{\pi_V} B[U_V, \sigma_i] + V + V \subset V + V + V \subset 4V,\end{aligned}$$

where the next to last inclusion follows by (1). Since $V$ is arbitrary, this completes the proof.

If condition B3 is strengthened so that $B[y, \sigma]$ is completely additive *uniformly* for $y\in U$, where $U$ is an arbitrary element of $\mathcal{V}$ (we say that $B[y, \sigma]$ is *uniformly completely additive*), then we can prove that boundedness[30] of $y(\sigma)$ implies $B$-summability and thus obtain the following theorem.

13.4. THEOREM. *Let $\mathfrak{X}$ be complete relative to $\mathcal{V}$ and let $B[y, \sigma]$ be uniformly*

---

[29] It is easy to prove, using B1–B4, that $B[y, \sigma_0]=\theta$ for every $y\in\mathfrak{Y}$ implies $B[y, \sigma]=\theta$ for every $\sigma\subset\sigma_0$ and $y\in\mathfrak{Y}$.

[30] The function $y(\sigma)$ is said to be *bounded* provided the set $y(M)$ is bounded.

*completely additive. Then, if $y(\sigma)$ is bounded, almost separable* (*B*) *and measurable* ($\mathfrak{M}$) *on the set $M \cap C\sigma_0$, it follows that $y(\sigma)$ is $\mathcal{V}_B$-integrable on each $\sigma$.*

## PART IV. RELATION TO OTHER INTEGRALS

14. **The Kolmogoroff integral.** The following theorem is a direct consequence of definitions; therefore the proof will be omitted.

14.1. THEOREM. *If $\mathfrak{X}$ is taken to be the real numbers, then the $\mathfrak{S}\mathcal{V}$-integral includes*[31] *the Kolmogoroff single-valued integral* [17, p. 663].

15. **The Phillips integral.** Consider the special "bilinear" functions $m(\sigma)x$, where $m(\sigma)$ is a completely additive positive measure function over $\mathfrak{M}$ and $x$ is an element of the convex linear topological space $\mathfrak{X}$.

15.1. LEMMA. *Let $X, Y \subset \mathfrak{X}$, $\sigma = \bigcup_{i=1}^n \sigma_i$, $\sigma_i \cap \sigma_j = 0$ $(i \neq j)$. Then $Y + m(\sigma)X \subset V_{cl}$, where $V \in \mathcal{V}$, implies $Y + \sum_{i=1}^n m(\sigma_i) X \subset V_{cl}$.*

We have $m(\sigma_i) Y + m(\sigma) m(\sigma_i) X \subset m(\sigma_i) V_{cl}$ $(i = 1, \cdots, n)$. Summing these relations over $i$ gives

$$\sum_{i=1}^n m(\sigma_i) Y + m(\sigma) \sum_{i=1}^n m(\sigma_i) X \subset \sum_{i=1}^n m(\sigma_i) V_{cl}.$$

But $V_{cl}$ is convex; therefore $\sum_{i=1}^n m(\sigma_i) V_{cl} = m(\sigma) V_{cl}$. Moreover $m(\sigma) Y \subset \sum_{i=1}^n m(\sigma_i) Y$; hence $Y + \sum_{i=1}^n m(\sigma_i) X \subset V_{cl}$, provided $m(\sigma) \neq 0$. Since the lemma is obviously true if $m(\sigma) = 0$, this completes the proof.

15.2. THEOREM. *The "bilinear" function $B[x, \sigma] = m(\sigma)x$ satisfies all of the conditions* B1–B5 (*including uniform complete additivity*); *therefore the entire theory of the $\mathcal{V}_B$-integral applies.*

All of the conditions are obviously satisfied except B4 which follows directly from Lemma 15.1. Observe that in the present case the constant of B4 can be taken as $1 + e$, where $e > 0$ is arbitrary.

Definition 10.1 of the $\mathcal{V}_B$-integral reduces in this case to precisely the definition used by Phillips [12, p. 118]; therefore

15.3. THEOREM. *For the case $B[x, \sigma] = m(\sigma)x$, the $\mathcal{V}_B$-integral reduces to the Phillips integral.*

16. **The Price integral.** Consider the special "bilinear" function $\tau(\sigma)x$, where $x$ is an element of a Banach space $\mathfrak{X}$ with its norm topology of spheres having center $\theta$ and where $\tau(\sigma)$ is a linear continuous transformation of $\mathfrak{X}$ into itself. $\tau(\sigma)x$ means the result of transforming $x$ by $\tau(\sigma)$. G. B. Price has

---

(31) One integral notion is said to include a second provided every function integrable according to the second notion is also integrable according to the first and to the same value.

defined an integral for this situation by first subjecting $\tau(\sigma)$ to the following conditions(32) [13, properties (8.1)–(8.3)].

T1. *If $\tau(\sigma)$ is the identically zero transformation, then $\sigma'\subseteq\sigma$ implies that $\tau(\sigma')$ is also identically zero.*

T2. *If $\tau(\sigma)$ is not the identically zero transformation, then it has a continuous inverse $\tau^{-1}(\sigma)$.*

T3. *For every sequence $\{\sigma_n\}$ of disjoint elements of $\mathfrak{M}$, $\tau(\bigcup\sigma_n)=\sum\tau(\sigma_n)$, where the series is unconditionally convergent according to the norm topology in the space of transformations.*

T4. *The generalized convex operator $C^*$ generated by $\tau(\sigma)$ is bounded* [13, pp. 7–10]. *The bound will be denoted by $\beta'$.*

16.1. Theorem. *If $\tau(\sigma)$ satisfies* T1–T4, *then the "bilinear" function $B[x, \sigma]=\tau(\sigma)x$ satisfies conditions* B1–B5 *(including uniform complete additivity). Therefore the entire theory of the $\mathcal{V}_B$-integral applies.*

All of the conditions are obviously satisfied except B4 for which we make the following proof.

Let $X_i\subset\mathfrak{X}$, $\sigma_i=\bigcup_{j=1}^{n_i}\sigma_i^j$, $\sigma_i\cap\sigma_j=0$ $(i\neq j)$, $\sigma_i^j\cap\sigma_i^k=0$ $(j\neq k)$, and assume $\sum_{i=1}^m\tau(\sigma_i)X_i\subset V_r$, where $V_r$ is a sphere of radius $r$ and center $\theta\in\mathfrak{X}$. The thing to be proved is that(33) $\sum_{i=1}^m\sum_{j=1}^{n_i}\tau(\sigma_i^j)X_i\subset\beta'(V_r)_{cl}$. In view of T1, we can evidently assume $\tau(\sigma_i)$ not identically zero $(i=1, \cdots, m)$. For simplicity let $\mu_i^j=\tau(\sigma_i^j)\tau^{-1}(\sigma_i)$, then

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\tau(\sigma_i^j)X_i=\sum_{i=1}^{m}\sum_{j=1}^{r_i}\mu_i^j\tau(\sigma_i)X_i.$$

Moreover, since $\sum_{j=1}^{n_i}\mu_i^j=I$, where $I$ is the identity transformation, we have the following

$$\sum_{i=1}^{m}\sum_{j=1}^{n_i}\mu_i^j\tau(\sigma_i)X_i\subset\sum_{j_1=1}^{n_i}\cdots\sum_{j_m=1}^{n_m}\mu_1^{j_1}\cdots\mu_m^{j_m}\left(\sum_{i=1}^{m}\tau(\sigma_i)X_i\right)$$
$$\subset\sum_{j_1=1}^{n_1}\cdots\sum_{j_m=1}^{n_m}\mu_1^{j_1}\cdots\mu_m^{j_m}V_r.$$

But $\{\mu_1^{j_1}\cdots\mu_m^{j_m}\}$ is obviously an element of the generalized convex operator $C^*$; therefore $\sum_{j=1}^m\sum_{j=1}^{n_i}\tau(\sigma_i^j)X_i\subset\beta'(V_r)_{cl}$.

The following theorem is an easy consequence of Theorems (11.4) and (4.11) of Price's paper.

(32) Condition T1 was not stated explicitly by Price but is implicit in the proof of part 6.9 of his Theorem 6.4. Also the situation discussed here is a bit more restricted than that considered by Price, since we require $\tau(\sigma)$ to be defined for every $\sigma\in\mathfrak{M}$ while Price admitted certain sets with "infinite measure" (see Footnote 14 above).

(33) Observe that the constant $\beta$ of condition B4 can then be taken as $\beta'+e$, where $e>0$ is arbitrary.

16.2. THEOREM. *If $B[x, \sigma]=\tau(\sigma)x$, then the $\mathcal{V}_B$-integral includes the Price integral.*

17. **Open questions.** Is it possible to dispense throughout with the condition that the space $\mathfrak{X}$ be convex?

Kolmogoroff [7] has also given a definition of a multi-valued integral for real functions. What is the precise relationship of this Kolmogoroff integral to the $\mathcal{V}$-integral when $\mathfrak{X}$ is the real numbers?

Is the specialization of the $\mathcal{V}_B$-integral which includes the Price integral actually equivalent to it?

Is condition T4 on $\tau(\sigma)$ equivalent, in the presence of T1–T3, to condition B4 on $B[x, \sigma]=\tau(\sigma)x$?

## BIBLIOGRAPHY

1. S. Banach, *Théorie des Opérations Linéaires*, Monografje Matematyczne, Warsaw, 1934.

2. Garrett Birkhoff, *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–56.

3. Nelson Dunford, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305–356.

4. L. M. Graves, *On the completing of a Hausdorff space*, Annals of Mathematics, (2), vol. 38 (1937), pp. 61–64.

5. F. Hausdorff, *Mengenlehre*, Berlin, 1927.

6. T. H. Hildebrandt, *On unconditional convergence in normed vector spaces*, Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 957–962.

7. A. Kolmogoroff, *Untersuchungen über den Integralbegriff*, Mathematische Annalen, vol. 103 (1930), pp. 654–696.

8. ———, *Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes*, Studia Mathematica, vol. 5 (1934), pp. 29–33.

9. K. Kunisawa, *Some theorems on abstractly-valued functions in an abstract space*, Proceedings of the Imperial Academy, Tokyo, vol. 16 (1940), pp. 68–72.

10. E. H. Moore and H. L. Smith, *A general theory of limits*, American Journal of Mathematics, vol. 44 (1922), pp. 102–121.

11. J. von Neumann, *On complete topological spaces*, these Transactions, vol. 37 (1935), pp. 1–20.

12. R. S. Phillips, *Integration in a convex linear topological space*, these Transactions, vol. 47 (1940), pp. 114–145.

13. G. B. Price, *The theory of integration*, these Transactions, vol. 47 (1940), pp. 1–50.

14. S. Saks, *Addition to the note on some functionals*, these Transactions, vol. 35 (1933), p. 967.

15. ———, *Theory of the Integral*, Monografje Matematyczne, Warsaw, 1937.

16. J. V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4 (1938), pp. 157–169.

UNIVERSITY OF MICHIGAN
ANN ARBOR, MICH.

# THE RESTRICTED PROBLEM OF THREE BODIES

BY
MONROE H. MARTIN

The difficulty of the problem of three bodies led Jacobi(1) to introduce a simplifying assumption, designed to make the problem more amenable to mathematical attack, but such that the problem retains its astronomical significance. In the restricted problem of three bodies Jacobi postulated that two masses known as *finite masses* revolve perpetually in concentric circles about their common center of mass in accordance with the laws of the two body problem and required the motion of a third mass termed the *infinitesimal mass* under the assumption that it is attracted by the two finite masses according to the Newtonian law of gravitation.

In this paper the scope of the problem is enlarged by permitting the two finite masses to move in accordance with an arbitrarily chosen solution of the two body problem. One is immediately led to three types of restricted problems according as one finite mass moves in an ellipse(2), parabola or hyperbola about the other as focus. As might be expected, considerable simplification occurs when the conic section degenerates into a line segment.

A restricted problem of three bodies may always be reduced to a *quasi-Lagrangian system*(3)

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_r} - \frac{\partial L}{\partial q_r} + k\frac{\partial L}{\partial \dot{q}_r} = 0, \qquad L = L(q_r, \dot{q}_s, t), \qquad k = k(t),$$

by the introduction of suitable variables. Such systems reduce to Lagrangian systems upon introduction of a Lagrangian function $\bar{L} = e^l L$, where $l = \int k dt$. Parts I, II, III of the paper are devoted to a study of these systems, partly with a view of their applications to the restricted problem of three bodies

Presented to the Society in two parts, the first under the title *Restricted problems in three bodies* on December 29, 1938 and the second under the title *Quasi-Lagrangian systems* on April 26, 1940; received by the editors June 27, 1941, and, in revised form, March 27, 1942.

(1) R. Marcolongo, *Il problema dei tre corpi*, Milan, 1919, p. 97, ascribes the problem to Jacobi. Recently the following papers, among others, have appeared on the problem: G. D. Birkhoff, *Sur le problème restreint des trois corps*, Annali della R. Scuola Normale Superiore di Pisa, (2), vol. 4 (1935), pp. 1–40 (first memoir) and (2), vol. 5 (1936), pp. 1–42 (second memoir); A. Wintner, *Beweis des E. Strömgrenschen dynamischen Abschlussprincips der periodischen Bahngruppen im restringierten Dreikörperproblem*, Mathematische Zeitschrift, vol. 34 (1931), pp. 321–349, where further references are given. See also A. Wintner, *The Analytical Foundations of Celestial Mechanics*, Princeton, 1941.

(2) The restricted problem of elliptic type has been investigated by F. R. Moulton, *Periodic Orbits*, Carnegie Institute of Washington Publication, 1920, pp. 217–284.

(3) Special cases of these systems have been considered by Elliott. See P. Appell *Traité de Mécanique Rationelle*, Paris, 1902, vol. 1, pp. 582–583, where further references are given.

and partly because of their intrinsic interest as examples of non-conservative systems. Among other things, their "limiting motions" are studied; in particular, conditions sufficient to insure that a motion tends toward equilibrium are obtained. For *quasi-conservative* ($L_t \equiv 0$, $k=$const.) systems it is possible to obtain a generalization of the energy integral and of the principle of least action, the latter involving a Mayer calculus of variations problem. Parts IV, V, VI deal with the restricted problem of three bodies. In the main they are concerned with the behavior of the infinitesimal mass as the two finite masses recede to infinity or approach (or leave) a collision along a degenerate conic section.

## I. Quasi-Lagrangian systems

1. **General principles.** A dynamical system with $n$ degrees of freedom and equations of motion([4])

$$(1)\qquad \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_r} - \frac{\partial L}{\partial q_r} + k\frac{\partial L}{\partial \dot{q}_r} = 0, \qquad L = (a_{rs}\dot{q}_r\dot{q}_s/2) + b_r\dot{q}_r + c, \qquad k = k(t),$$

is termed a *quasi-Lagrangian system*. If $k$ is a positive constant, the quasi-Lagrangian system becomes a dissipative system for which Rayleigh's dissipation function([5]) is proportional to the Lagrangian function. It is readily verified that the equations of motion may be given the variational form

$$(2)\qquad \delta\int_{t_0}^{t_1} e^l L dt = 0, \qquad l = \int k dt,$$

there being no variation in the time $t$, nor in the end points of the varied curves. Subjected to Legendre's transformation they take the canonical form([6])

$$(3)\qquad \dot{q}_r = H_{p_r}, \qquad \dot{p}_r = -H_{q_r} - kp_r, \qquad H = (a^{rs}2/)(p_r - b_r)(p_s - b_s) - c,$$

the integration of which is equivalent to the determination of a complete solution of the partial differential equation

$$(4)\qquad S_t + kS + H(t, q_r, S_{q_r}) = 0.$$

This partial differential equation accordingly plays the role of the Hamilton-Jacobi partial differential equation.

In place of Liouville's theorem which likens the flow in the phase space

---

([4]) All functions are assumed analytic in their arguments. The dot denotes differentiation with respect to the time $t$. The matrix $\|a_{rs}\|$ is assumed to be positive definite and the repeated indices denote summation from 1 to $n$.

([5]) For the theory of Rayleigh's dissipation functions see E. T. Whittacker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, Cambridge, 1927, pp. 230–231.

([6]) The reciprocal matrix of $\|a_{rs}\|$ is denoted by $\|a^{rs}\|$.

of a Lagrangian system to that of an incompressible fluid, the flow in the phase space of a quasi-Lagrangian system obeys the law $Ve^{nt}=\text{const.}$, where $V$ is the $2n$-dimensional volume of a portion of the phase space at time $t$. To prove this, one notes that the condition(7) for $\int M dV$ to be an integral invariant of (3) yields $Me^{-nt}=\text{const.}$

The rate at which the energy $H$ is changing along a motion of (3) is expressed by

$$dH/dt = -kp_r H_{p_r} + H_t,$$

and the system is termed *acquisitive* or *dissipative* in a certain time interval according as $dH/dt \gtrless 0$ holds in this time interval for every motion of the system. The two forms of $H$ important for the restricted problem of three bodies turn out to be

$$\begin{aligned} H &= (a^{rs}/2)(p_r - e^{-t}b_r)(p_s - e^{-t}b_s) - c, \\ H &= (a^{rs}/2)(p_r - e^{-t}b_r)(p_s - e^{-t}b_s) - e^{-t}c, \end{aligned} \tag{5}$$

where $a^{rs}$, $b_r$, $c$ are functions of $q_r$ only. One finds

$$\begin{aligned} dH/dt &= -ka^{rs}(p_r - e^{-t}b_r)(p_s - e^{-t}b_s), \\ d(e^tH)/dt &= -(k/2)e^t a^{rs}(p_r - e^{-t}b_r)(p_s - e^{-t}b_s), \end{aligned}$$

holding, respectively, so that in both cases the system is acquisitive or dissipative according as $k \lessgtr 0$.

2. **Limiting motions.** A motion $\overline{M}$ defined(8) for $a<t<b$ is an *ω-limiting motion* of a motion $M$ defined for $t>t_0$ if, given any subinterval $a_0 \leqq t \leqq b_0$ of $a<t<b$, an arbitrarily fixed, small positive number $\delta$, and an arbitrarily fixed, large positive number $T$, there exists a $\tau>T$ such that the point $P(t)$ on $M$ is at a distance less than $\delta$ from the point $\overline{P}(t-\tau)$ on $\overline{M}$ for $a_0 \leqq t-\tau \leqq b_0$. An *ω-limiting point* of $M$ is a point of accumulation of a sequence of points $P(t_i)$ on $M$ for which $t_i \to +\infty$. α-limiting motions and α-limiting points are defined similarly for $t \to -\infty$. Assuming that the second members in (3) are regular analytic in $q_r$, $p_s$, $t$ throughout the phase space with the exception of the points of a set $S$ at which singularities occur for $t>t_0$, a motion defined for $t>t_0$ is *positively stable* if it remains in a bounded portion of the phase space and does not come arbitrarily close to $S$, otherwise $M$ is *positively instable.* Negatively stable (instable) motions are defined similarly and a motion both positively and negatively stable is termed *stable.*

If $H$, $H_{q_r}$, $H_{p_s}$ tend towards limiting functions $\overline{H}$, $\overline{H}_{q_r}$, $\overline{H}_{p_s}$ uniformly in any bounded, closed subregion of the phase space not containing points of $S$ while $k$ tends toward a finite limit $\bar{k}$ as $t \to +\infty$, the ω-limiting motions of a

(7) See E. T. Whittacker, op. cit., pp. 283–284.

(8) The possibilities $a=-\infty$, $b=+\infty$ are not to be excluded.

motion $M$ of (3) are(9) motions of the limiting system

$$\dot{q}_r = \bar{H}_{p_r}, \qquad \dot{p}_r = -\bar{H}_{q_r} - \bar{k}p_r. \tag{6}$$

In case $M$ is positively stable its $\omega$-limiting motions comprise a set of stable motions approached uniformly by $M$ as $t \to +\infty$.

THEOREM 1. *If the Hamiltonian function $H$ has either form in* (5) *and if $k$ tends toward a finite positive limit $\bar{k}$ as $t \to +\infty$ the $\omega$-limiting motions of a positively stable motion $M$ of* (3) *are equilibrium motions*(9) *of the system* (6) *with* $\bar{H} = (a^{rs}p_r p_s/2) - c$, *or* $\bar{H} = a^{rs}p_r p_s/2$.

Along $M$ the energy $H$ is eventually a monotone decreasing function of $t$ and will accordingly approach a finite limit $\bar{h}$, since $M$ is positively stable. Along an $\omega$-limiting motion $\bar{M}$ of $M$ we have $\bar{H} = \bar{h}$. To see this let $\bar{P}$ be any point of $\bar{M}$ and $P$ be the point on $M$ at time $t$. Clearly

$$|\bar{H}(\bar{P}) - \bar{h}| \leq |\bar{H}(\bar{P}) - \bar{H}(P)| + |\bar{H}(P) - H(P)| + |H(P) - \bar{h}|.$$

Let $\epsilon$ be an arbitrarily small positive number. In view of the uniform convergence of $H$ towards $\bar{H}$ we may select $t_1$ such that $|\bar{H}(P) - H(P)| < \epsilon$ for all $t > t_1$. Since $H(P)$ tends to $\bar{h}$ as a limit, there exists a $t_2$ for which $|H(P) - \bar{h}| < \epsilon$ for all $t > t_2$. Finally $\bar{H}(P)$ is a continuous function of $P$ and $\bar{P}$ is an $\omega$-limiting point of $M$. Hence there exists a value of $t$ greater than either $t_1$ or $t_2$ for which $|\bar{H}(\bar{P}) - \bar{H}(P)| < \epsilon$. It follows that $|\bar{H}(\bar{P}) - \bar{h}|$ is arbitrarily small and therefore equals zero.

Along $\bar{M}$ we have $d\bar{H}/dt = -\bar{k}a^{rs}p_r p_s = 0$ and therefore, since $\|a^{rs}\|$ is positive definite, the $p_r$ are zero on $\bar{M}$. Hence $\bar{M}$ is an equilibrium motion.

It is obvious that these systems possess no periodic motions.

## II. QUASI-CONSERVATIVE SYSTEMS

1. **General principles.** A quasi-Lagrangian system is termed *quasi-conservative* if $L_t \equiv 0$ and $k$ is a constant other than zero. For such systems there exists a generalization of the energy integral

THEOREM 2. *If $S$ be defined along a motion $M$ by*

$$e^{kt}S = e^{kt}S_0 + \int_{t_0}^{t} e^{kt}(p_r H_{p_r} - H)dt, \tag{7}$$

*the quantity $(H + kS)e^{kt}$ retains a constant value along $M$.*

Placing $z = S$, $p = S_t$, $p_r = S_{q_r}$ in (4), the differential equations of the char-

(9) For the treatment of steady flows, see G. D. Birkhoff, *Quelques théorèmes sur le mouvement des systèmes dynamiques*, Bulletin de la Société Mathématique de France, vol. 40 (1912), pp. 305–323. For a treatment of dissipative systems not involving the time explicitly, see his book *Dynamical Systems*, American Mathematical Society Colloquium Publications, vol. 9, 1927, pp. 31–32.

acteristic strips of (4) are

$$(8)\qquad t' = 1,\quad q_r' = H_{p_r},\quad p' = -kp,\quad p_r' = -H_{q_r} - kp_r,\quad z' = p + p_r H_{p_r},\qquad (' = d/du).$$

They possess the integral

$$(9)\qquad p + kz + H(q_r, p_s) = \text{const.}$$

By adjoining

$$t = u,\qquad p = p_0 e^{-k(t-t_0)},\qquad z = z_0 + \int_{t_0}^{t} (p_r H_{p_r} + p)dt,$$

to the equations of a motion of (3), one obtains a solution of (8). Taking $p_0$ arbitrarily, $z_0$ is determined so that the constant in (9) equals zero, and the relation obtained is used to eliminate $p$ from the last equation in (8) to obtain

$$dz/dt + kz = p_r H_{p_r} - H,$$

which, when integrated, yields (7). The theorem then follows from (9).

COROLLARY. *If $H$ is homogeneous of degree two in $q_r$, $p_s$, a quasi-conservative system has a first integral* $(H+(k/2)p_r q_r)e^{kt} = \text{const}$.

To establish the corollary, set $2H = q_r H_{q_r} + p_r H_{p_r}$ in (7) and substitute for $H_{q_r}$, $H_{p_r}$ from (3) to obtain $2e^{kt}S = e^{kt}p_r q_r + \text{const}$.

The generalization of the principle of least action to quasi-conservative systems rests on the following lemma[10].

LEMMA. *The projections of the characteristic curves of the partial differential equation*

$$f(x_r, z, p_s) = 0,\qquad p_s = z_{x_s},\ f_{p_1}^2 + \cdots + f_{p_n}^2 \neq 0,\ \det \|f_{p_r p_s}\| \neq 0,$$

*upon the space of the independent variables $x_r$ are the extremals of the Mayer calculus of variations problem $\delta z_1 = 0$, given that*

$$dz/dt = F(x_r, z, \dot{x}_s),\qquad \delta \overset{0}{x_r} = \delta \overset{0}{z} = 0,\qquad \delta \overset{1}{x_r} = 0,$$

*where $\zeta - z = F(x_r, z, x_s - \xi_s)$ is the equation of the Monge cone for the partial differential equation at the point $x_r$, $z$.*

THE PRINCIPLE OF LEAST ACTION. *The projections of the characteristic curves of the partial differential equation $kz + H(q_r, p_s) = 0$, $p_s = z_{q_s}$ upon the space of the coordinates $q_r$ are the extremals of the Mayer calculus of variations problem $\delta z_1 = 0$, given that*

$$\dot{z} = (2(c - kz)a_{rs}\dot{q}_r\dot{q}_s)^{1/2} + b_r\dot{q}_r,\qquad \delta \overset{0}{q_r} = \delta \overset{0}{z} = 0,\qquad \delta \overset{1}{q_r} = 0.$$

(10) For the proof of this lemma for the partial differential equation $f(x, y, p, q) = 0$, see A. Kneser, *Lehrbuch der Variationsrechnung*, Braunschweig, 1925, pp. 157–160.

To arrive at the principle, we set $S=he^{-kt}+W$, where $h=$const. and $W=W(q_1, \cdots, q_n)$ in (4) to obtain the partial differential equation $kW+H(q_r, W_{q_s})=0$. Placing $z=W$, $p_s=W_{q_s}$ and calculating the equation of the Monge cone for this partial differential equation, the truth of the principle follows from the lemma.

To see why the principle is to be regarded as a generalization of the principle of least action for conservative systems, it will be recalled that placing $S=-ht+W$ in the Hamilton-Jacobi partial differential equation of a conservative system leads to the partial differential equation $-h+H(q_r, W_{q_s})=0$. The projections upon the space of the coordinates $q_r$ of the characteristics of this partial differential equation are the extremals of the Mayer calculus of variations problem $\delta z^1=0$, given that

$$\dot{z} = (2(c + h)a_{rs}\dot{q}_r\dot{q}_s)^{1/2} + b_r\dot{q}_r, \qquad \delta \overset{0}{q_r} = \delta \overset{0}{z} = 0, \qquad \delta \overset{1}{q_r} = 0.$$

When this is formulated as an ordinary calculus of variations problem with fixed end points

$$\delta \int_{t_0}^{t_1} \{(2(c + h)a_{rs}\dot{q}_r\dot{q}_s)^{1/2} + b_r\dot{q}_r\} dt = 0,$$

one obtains the classic expression for the principle of least action.

2. **Characteristic exponents**(11). Assuming that the origin of the phase space is an equilibrium motion of (3), the equations of variation written in matrix form are

$$dx/dt = Ax, \qquad A = GH - kK. \tag{10}$$

Here $x$ denotes a matrix with $2n$ rows and one column, the first and last $n$ rows being occupied by the variations of $q_r$, $p_s$, respectively, while(12)

$$G = \left\| \begin{matrix} \omega & \varepsilon \\ -\varepsilon & \omega \end{matrix} \right\|, \quad H = \left\| \begin{matrix} H_{q_r q_s} & H_{q_r p_s} \\ H_{p_r q_s} & H_{p_r p_s} \end{matrix} \right\|, \quad K = \left\| \begin{matrix} \omega & \omega \\ \omega & \varepsilon \end{matrix} \right\|, \quad M = \left\| \begin{matrix} \omega & \varepsilon \\ \varepsilon & \omega \end{matrix} \right\|,$$

where $\omega$, $\varepsilon$ denote the $n \times n$ zero and unit matrices, respectively, and the partial derivatives are evaluated at the origin. Denoting the transposed matrix by affixing a prime, one verifies that

$$\begin{aligned} &G' = -G, \qquad GG' = G'G = -G^2 = E; \\ &H' = H, \qquad K' = K, \qquad GKG = K - E, \end{aligned} \tag{11}$$

where $E$ is the unit matrix of $2n$ rows and columns.

---

(11) For the theorems on characteristic exponents of conservative systems, see G. D. Birkhoff, *Dynamical Systems*, loc. cit., pp. 74–78, and A. Wintner, *Three notes on characteristic exponents and equations of variation in celestial mechanics*, American Journal of Mathematics, vol. 53 (1931), p. 609.

(12) The matrix $M$ defined here will not be needed until Theorem 4.

THEOREM 3. *The characteristic exponents may be divided into pairs in which the two members are of equal multiplicity and have $-k$ for their sum.*

The characteristic exponents are the roots of the equation $f(\lambda) = \det(A - \lambda E) = 0$. Replacing $A$ by its transposed, multiplying before and behind by $G$, it follows from (11) that $f(\lambda) = f(-\lambda - k)$, and the theorem is proved.

It may be noted in passing that at least half of the characteristic exponents have negative (positive) real parts if $k > 0$ ($k < 0$).

THEOREM 4. *If the matrix $H + (k/2)M$ is definite, the characteristic exponents lie on the line $\Re(\lambda) = -k/2$ with $\lambda = -k/2$ excluded.*

Corresponding to a characteristic exponent $\lambda$, there is a solution of (10) given by $x = ae^{\lambda t}$, where $a$ is a constant matrix satisfying $(A - \lambda E)a = 0$. Multiplying on the left by $a'G$ and observing that $2GK = M + G$, it is found that

$$a'(H + (k/2)M)a = 0, \tag{12}$$

inasmuch as $G$ is skew-symmetric. If $\lambda$ is real, $a$ may be taken real and the assumption that $H + (k/2)M$ is definite is contradicted by (12). Thus no characteristic exponent is real.

We may regard (10) as the canonical equations of a quasi-conservative system whose Hamiltonian function $H$ is homogeneous of degree 2 in its arguments and write the first integral, previously obtained in the corollary to Theorem 1, in the matrix form $x'(H + (k/2)M)xe^{kt} = \text{const}$. For complex $\lambda$, the real solution $x + \bar{x}$ is formed and inserted in the energy integral. Keeping (12) in mind, one finds that $a'(H + (k/2)M)\bar{a} \exp(k + \lambda + \bar{\lambda})t = \text{const.}$, which requires $\Re(\lambda) = -k/2$.

## III. NATURAL QUASI-CONSERVATIVE SYSTEMS

A quasi-Lagrangian system (3) is a *natural* system if $b_r = 0$. Taking $b_r = 0$ in (5), it follows from Theorem 1 that the $\omega$-limiting motions of a positively stable motion of a natural quasi-conservative system (3) are equilibrium motions of (3). Such systems accordingly possess no periodic motions.

THEOREM 5. *To each characteristic constant $\delta_r$ of the matrix $\|-c_{q_r q_s}\|$, evaluated at an equilibrium motion of* (3), *there correspond two characteristic exponents given by the roots of the equation*

$$\lambda^2 + k\lambda + \delta_r = 0.$$

Referring to §2 of II, it is apparent that

$$H = \begin{Vmatrix} \beta & \omega \\ \omega & \alpha \end{Vmatrix}, \qquad \alpha = \|a^{rs}\|, \qquad \beta = \|-c_{q_r q_s}\|,$$

and it may be verified that the transformation $x = Ty$, where

$$T = \begin{Vmatrix} \gamma & \omega \\ \omega & \gamma'^{-1} \end{Vmatrix}, \qquad T'HT = \begin{Vmatrix} \gamma'\beta\gamma & \omega \\ \omega & \gamma^{-1}\alpha\gamma'^{-1} \end{Vmatrix},$$

carries (10) into $dy/dt = By$, with $B = GT'HT - kK$. Choosing $\gamma$ so that $\gamma'\alpha^{-1}\gamma = \epsilon$, $\gamma'\beta\gamma = \delta$ where $\delta$ is a diagonal matrix with the characteristic constants of $\beta$ along the principal diagonal, the theorem follows inasmuch as the characteristic exponents satisfy the equation det $(B - \lambda E) = 0$.

It is clear that there are no purely imaginary characteristic exponents. If the characteristic exponents $\lambda_1, \cdots, \lambda_m$ have negative real parts and the remaining $2n-m$ have positive real parts, the equilibrium motion is of *negative* or *positive general type*, according as none of the linear commensurability relations

$$(13)\qquad \begin{aligned} &\mathrm{I}\begin{cases} p_1\lambda_1 + \cdots + p_m\lambda_m = \lambda_j, & j = 1, \cdots, m; \\ p_1 + \cdots + p_m \geqq 2, & p_j = 0, 1, \cdots; \end{cases} \\ &\mathrm{II}\begin{cases} p_{m+1}\lambda_{m+1} + \cdots + p_{2m}\lambda_{2m} = \lambda_k, & k = m+1, \cdots, 2n; \\ p_{m+1} + \cdots + p_{2m} \geqq 2, & p_k = 0, 1, \cdots; \end{cases} \end{aligned}$$

in I or II is satisfied.

THEOREM 6. *If an equilibrium motion $M_0$ is of negative [positive] general type, a suitably restricted neighborhood of $M_0$ contains an analytic m-dimensional [(2n−m)-dimensional] surface, the points of which tend towards $M_0$ as a limit as $t \to +\infty$ $[-\infty]$. No other points of this neighborhood tend towards $M_0$ as a limit as $t \to +\infty$ $[-\infty]$. The number m is the number of characteristic exponents of $M_0$ with negative real parts*(13).

## IV. THE RESTRICTED PROBLEM OF THREE BODIES

Two finite masses $\mu$ and $1-\mu$ move in a fixed plane under their mutual gravitational attraction. The *restricted problem of three bodies* is the problem of determining the motion in the fixed plane of a third mass (the infinitesimal mass) subject to the gravitational attraction of the two finite masses, the motion of which is assumed to proceed independently of its presence. The problem is said to be of *elliptic, parabolic* or *hyperbolic type* according as the orbit of $\mu$ about $1-\mu$ is an ellipse, parabola or hyperbola. When the conic section degenerates into a line segment the problem is termed *rectilinear*, otherwise we term the problem *general*.

If we take the center of mass of the two finite masses as origin of a rectangular coordinate system $(\xi, \eta)$ having a fixed orientation in the plane of motion of the two finite masses, it is known that

$$(14)\qquad \rho'' - \rho\theta'^2 = -\kappa^2\rho^{-2}, \quad \rho^2\theta' = \alpha = \text{const.}, \quad (\kappa^2 = \text{gravitational constant}),$$

(13) A proof of this theorem has been given by the writer in Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 475–481. See also C. L. Siegel, *Der Dreierstoss*, Annals of Mathematics, (2), vol. 42 (1941), pp. 156–165.

where $\rho$, $\theta$, respectively, denote the length and angle of inclination with the positive $\xi$-axis of the line joining the two finite masses, differentiation with respect to the time $\tau$ being indicated by a prime. Denoting the distances of the infinitesimal mass from $1-\mu$, $\mu$ by $\rho_1$, $\rho_2$, respectively, the Lagrangian function $\Lambda$ for its motion is

$$\Lambda = (\xi'^2 + \eta'^2)/2 + \kappa^2((1-\mu)/\rho_1 + \mu/\rho_2).$$

THEOREM 7. *The introduction of new variables $x$, $y$, $t$ by*

$$\xi + i\eta = \rho e^{i\theta}(x + iy), \qquad \kappa d\tau = \rho^{3/2}dt, \qquad \rho^{3/2} > 0, \tag{15}$$

*reduces the restricted problem of three bodies to a quasi-Lagrangian system in which*

$$\begin{aligned} L &= (\dot{x}^2 + \dot{y}^2)/2 + be^{-l}(x\dot{y} - y\dot{x}) + \Omega, & b &= \alpha/\kappa, \\ H &= [(p + be^{-l}y)^2 + (q - be^{-l}x)^2]/2 - \Omega, & l &= (\log \rho)/2, \end{aligned} \tag{16}$$

*where*

$$\begin{aligned} \Omega &= (x^2 + y^2)/2 + (1-\mu)/r_1 + \mu/r_2, \\ r_1^2 &= (x+\mu)^2 + y^2, \qquad r_2^2 = (x - 1 + \mu)^2 + y^2, \end{aligned} \tag{17}$$

*and differentiation with respect to $t$ is indicated by a dot. The system is acquisitive or dissipative according as the two finite masses approach or leave each other and reduces to a natural system for rectilinear problems.*

It may be verified that

$$\begin{aligned} \xi'^2 + \eta'^2 &= \rho^2(x'^2 + y'^2) + 2\rho^2\theta'(xy' - yx') \\ &\quad + (\rho'^2 + \rho^2\theta'^2)(x^2 + y^2) + 2\rho\rho'(xx' + yy'), \\ [\rho\rho'(x^2 + y^2)]' &= (\rho'^2 + \rho\rho'')(x^2 + y^2) + 2\rho\rho'(xx' + yy'). \end{aligned}$$

When these equations are subtracted with (14) in mind, the result substituted in $\Lambda$, and the term $[\rho\rho'(x^2+y^2)]'$ suppressed([14]), it is found that

$$\Lambda = (\rho^2/2)(x'^2 + y'^2) + \alpha(xy' - yx') + (\kappa^2/\rho)\Omega,$$

provided one observes that $\rho_1 = \rho r_1$, $\rho_2 = \rho r_2$.

Writing the equations of motion for the infinitesimal mass in the variational form $\delta\int_{\tau_0}^{\tau_1}\Lambda d\tau = 0$ and introducing the new independent variable $t$ defined in (15), this variational equation takes the form (2) with $l$, $L$ defined as in (16).

It will be observed that the Hamiltonian function $H$ has the first form in (5). The system is therefore acquisitive or dissipative as stated.

([14]) This may be done since the variation of its integral vanishes.

REMARK. *If the new independent variable is defined by $\kappa d\tau = \rho d\bar{l}$ the restricted problem of three bodies reduces to a quasi-Lagrangian system with*

$$(18)\qquad \begin{aligned} L &= (\dot{x}^2 + \dot{y}^2)/2 + be^{-l}(x\dot{y} - y\dot{x}) + e^{-l}\Omega, & b &= \alpha/\kappa, \\ H &= [(p + be^{-l}y)^2 + (q - be^{-l}x)^2]/2 - e^{-l}\Omega, & l &= \log \rho, \end{aligned}$$

*where the dot denotes differentiation with respect to $\bar{l}$. The Hamiltonian function has the second form in* (5), *the system being acquisitive or dissipative according as the finite masses approach or leave each other, and is a natural system for rectilinear problems.*

## V. THE RESTRICTED PROBLEM OF PARABOLIC TYPE

1. **The differential equations of motion.** If $\mu$ moves in a parabolic orbit about $1-\mu$, it is known[15] that

$$(19)\qquad \rho = (p/2)\sec^2(\theta/2), \quad 2\kappa\tau = p^{3/2}((1/3)\tan^3(\theta/2) + \tan(\theta/2)), \quad p > 0,$$

where $p$ is the semi-latus rectum of the parabolic orbit. In case $p=0$, the parabolic orbit degenerates into a line segment, and we have

$$(20)\qquad \rho = (3\kappa\tau/2^{1/2})^{2/3}.$$

When the variable $t$ of Theorem 7 is introduced in place of $\tau$, these equations are replaced by

$$(21)\qquad \rho = (p/2)\cosh^2(t/2^{1/2}), \qquad \sin(\theta/2) = \tanh(t/2^{1/2}),$$

$$(22)\qquad \rho = e^{\pm 2^{1/2} t}.$$

Thus, as $t$ ranges from $-\infty$ to $+\infty$, the mass $\mu$ describes the complete parabolic orbit ($p>0$) or describes the line segment ($p=0$), approaching a collision with $1-\mu$ as $t\rightarrow\pm\infty$ according as the negative or positive sign is taken in (22). Corresponding to (21), (22) the function $l$ in (16) is given by

$$l = (1/2)\log(p/2) + \log\cosh(t/2^{1/2}), \qquad l = \pm t/2^{1/2},$$

from which the differential equations (1) are found, on setting $c = b(2/p)^{1/2}$

$$(23)\qquad \begin{aligned} &\ddot{x} + (1/2^{1/2})\tanh(t/2^{1/2})\cdot\dot{x} - 2c\,\mathrm{sech}(t/2^{1/2})\cdot\dot{y} = \Omega_x, \\ &\ddot{y} + (1/2^{1/2})\tanh(t/2^{1/2})\cdot\dot{y} + 2c\,\mathrm{sech}(t/2^{1/2})\cdot\dot{x} = \Omega_y, \end{aligned}$$

$$(24)\qquad \ddot{x} \pm \dot{x}/2^{1/2} = \Omega_x, \qquad \ddot{y} \pm \dot{y}/2^{1/2} = \Omega_y,$$

the latter equations holding for the rectilinear problem, with the positive or negative sign taken according as the finite masses leave or approach each other. The canonical form (3) of (23) is

$$(25)\qquad \dot{x} = H_p, \qquad \dot{y} = H_q, \qquad \dot{p} = -H_x - kp, \qquad \dot{q} = -H_y - kq,$$

[15] See, for example, F. R. Moulton, *An Introduction to Celestial Mechanics*, New York, 1935, pp. 155–159.

where $k=(1/2^{1/2})\tanh(t/2^{1/2})$ and $H$ is defined in (16), while that of (24) is

$$\dot{x} = \overline{H}_p, \quad \dot{y} = \overline{H}_q, \quad \dot{p} = -\overline{H}_x - \bar{k}p, \quad \dot{q} = -\overline{H}_y - \bar{k}q, \tag{26}$$
$$\overline{H} = (p^2+q^2)/2 - \Omega, \qquad \bar{k} = \pm 1/2^{1/2}.$$

Corresponding to the positive and negative values for $\bar{k}$ we have two kinds of rectilinear problems; the former is termed the ***dissipative rectilinear problem*** and the latter the ***acquisitive rectilinear problem.***

2. **The flow in the phase space.** A comparison of (25) and (26) shows that the $\alpha$ and $\omega$-limiting motions in the general problem are, respectively, motions of the acquisitive and dissipative rectilinear problems.

The set $S$ of singularities of the system (25) comprises the points of the two-dimensional surfaces $x=-\mu$, $y=0$; $x=1-\mu$, $y=0$ corresponding to collisions of the infinitesimal mass with one or the other of the finite masses.

THEOREM 8. *If the energy $H$ of a motion tends to $-\infty$ as $t\to+\infty$ the motion is positively instable and the point $P$ in the $(x, y)$-plane corresponding to the infinitesimal mass either tends uniformly towards one of the points corresponding to the finite masses or else tends uniformly towards the point at infinity as $t\to+\infty$.*

Suppose that $P$ tends uniformly to neither of the points corresponding to the finite masses nor to the point at infinity as $t\to+\infty$. There would exist a positive $\epsilon$ and an infinite sequence $\{t_n\}$ of $t$-values for which $P$ would lie in the region

$$(x+\mu)^2+y^2>\epsilon, \qquad (x-1+\mu)^2+y^2>\epsilon, \qquad x^2+y^2>1/\epsilon,$$

and from the definitions of $H$, $\overline{H}$ in (16), (26) there would exist a constant $K$ such that $H(t_n)>K$, thus contradicting the hypothesis.

The equilibrium motions of the rectilinear problems are characterized by

$$p=0, \qquad q=0, \qquad \Omega_x=0, \qquad \Omega_y=0,$$

the latter two equations being satisfied at exactly five points $L_i$ in the $(x, y)$-plane known as libration points. $L_1$, $L_2$, $L_3$ lie on the $x$-axis with $L_1$ between, $L_2$ to the right of, and $L_3$ to the left of the points corresponding to the finite masses which in turn form equilateral triangles with $L_4$, $L_5$. Corresponding to $L_i$ there are five equilibrium motions $E_i$ in the phase space.

THEOREM 9. *If the point $P$ representing the infinitesimal mass remains in a bounded closed region of the $(x, y)$-plane containing $L_i$ and not containing the points corresponding to the finite masses, it tends uniformly towards a definite $L_i$ as $t\to\pm\infty$.*

In view of the restriction upon $P$ it is clear that $-\Omega$ has a finite lower bound. Since $H$ decreases monotonely for sufficiently large $t$, it follows that

$p^2+q^2$ remains bounded for $t>t_0$. The corresponding motion $M$ in the phase space is accordingly stable and it follows from Theorem 1 that the $\omega$-limiting motions of $M$ are the equilibrium motions $E_i$ of (26). Now $M$ approaches its set of $\omega$-limiting motions uniformly and therefore it approaches a definite $E_i$ uniformly as $t\rightarrow+\infty$.

The proof of the theorem has been given only for the general problem. It is evident that it remains in force for the rectilinear problem and when $t\rightarrow-\infty$.

Since a stable motion has one $E_i$ as a unique $\alpha$-limiting point and one as a unique $\omega$-limiting point, the stable motions may be divided into twenty-five classes, inasmuch as the $\alpha$- and $\omega$-limiting points conceivably may be chosen at random from the five $E_i$.

3. **The rectilinear problem.** The differential equations (24) for the acquisitive and dissipative rectilinear problems interchange when $t$ is replaced by $-t$. It will be sufficient, therefore, to consider one of these problems. We shall select the dissipative problem for further investigation.

THEOREM 10. *If stable motions other than equilibrium motions exist in the dissipative rectilinear problem, they fall into the following nine classes:*

(i) *$\alpha$-limiting point $E_4$ or $E_5$,* *$\omega$-limiting point $E_1$, $E_2$, $E_3$;*
(ii) *$\alpha$-limiting point $E_3$,* *$\omega$-limiting point $E_1$, $E_2$;*
(iii) *$\alpha$-limiting point $E_2$,* *$\omega$-limiting point $E_1$;*

*provided* $0<\mu<1/2$. *When* $\mu=1/2$ *there is no motion in* (ii) *with $E_2$ as $\omega$-limit point and when* $1/2<\mu<1$, *the roles of $E_2$, $E_3$ are interchanged.*

Along a motion $M$ of (26) not an equilibrium motion $\overline{H}$ decreases monotonely with increasing $t$ for all $t$. Therefore an $E_i$ cannot serve simultaneously as $\alpha$- and $\omega$-limiting point for $M$, since this would require $\overline{H}$ constant along $M$ and $d\overline{H}/dt=-(p^2+q^2)/2^{1/2}\equiv 0$ would imply that $M$ is an equilibrium motion, contrary to hypothesis.

If $\Omega_i$ is the value of $\Omega$ at $L_i$, it is known([16]) that $\Omega_2$ is greater than, equal to, or less than $\Omega_3$ according as $\mu$ is less than, equal to, or greater than $1/2$ and that $\Omega_1$ exceeds while $\Omega_4=\Omega_5$ is less than either of $\Omega_2$, $\Omega_3$. Therefore, if $\overline{H}_i$ indicates the value of $\overline{H}$ at $E_i$ it follows that $\overline{H}_2$ is less than, equal to, or greater than $\overline{H}_3$ according as $\mu$ is less than, equal to, or greater than $1/2$ and that $\overline{H}_1$ is less than while $\overline{H}_4=\overline{H}_5$ is greater than either $\overline{H}_2$ or $\overline{H}_3$.

It is therefore impossible for a stable motion to have one $E_4$, $E_5$ as $\alpha$-limiting point and the other for $\omega$-limiting point and the theorem is an immediate consequence of the above inequalities and the monotone character of $\overline{H}$.

A study of the nature of the flow in the phase space in the neighborhoods of $E_i$ leads to a sharper classification of stable motions. Prior to such study a lemma dealing with the characteristic exponents of the $E_i$ is needed.

([16]) See A. Wintner, *The Analytical Foundations of Celestial Mechanics*, Princeton, 1941, pp. 364–366.

LEMMA. *There is one positive and one negative characteristic exponent for $E_1$; the two remaining are conjugate complex numbers with negative real parts. There is one positive characteristic exponent for $E_2$; the remaining three have negative real parts and there exists a constant $\mu^*$ such that of these three: one is real and the other two are conjugate complex numbers for $0<\mu<\mu^*$; all are real with two equal to $-2^{-3/2}$ and differing from the third if $\mu=\mu^*$, all are real and different for $\mu^*<\mu<1$. At $E_3$ the situation is analogous to that at $E_2$ and at $E_4$, $E_5$ there are two distinct positive and two distinct negative characteristic exponents.*

Consider the three libration points $L_i$ $(a_i, 0)$ on the $x$-axis. One finds that

$$\Omega_{xx}(a_i, 0) = 1 + 2A_i, \qquad \Omega_{xy}(a_i, 0) = 0, \qquad \Omega_{yy} = 1 - A_i, \tag{27}$$

where

$$A_i = \frac{1-\mu}{|a_i+\mu|^3} + \frac{\mu}{|a_i+\mu-1|^3}.$$

If we use (27) to compute the characteristic constants of the matrix $\|-c_{q_r q_s}\|$ in Theorem 5, and take $k=1/2^{1/2}$ in this theorem, the characteristic exponents $\lambda$ of $E_i$ turn out to be

$$\lambda = 2^{-3/2}(-1 \pm (9+16A_i)^{1/2}), \qquad \lambda = 2^{-3/2}(-1 \pm (9-8A_i)^{1/2}),$$

from which the statements of the theorem relative to the characteristic exponents at $E_1$, $E_2$, $E_3$ follow, inasmuch as it may be proved that $A_1>4$ and $A_2$ $(A_3)$ decreases (increases) monotonely and continuously from 4 (1) to 1 (4) as $\mu$ varies from 0 to 1.

After computing the partial derivatives of second order for $\Omega$ at $L_4$, $L_5$ and obtaining the characteristic constants of $\|-c_{q_r q_s}\|$, it is found that the characteristic exponents $\lambda$ of $E_4$, $E_5$ are given by

$$\lambda = 2^{-3/2}(-1 \pm (13 \pm 12(1-3\mu(1-\mu))^{1/2})^{1/2}),$$

from which follow the properties for the characteristic exponents of $E_4$, $E_5$.

In view of this lemma, Theorem 6 ($n=2$, $m=3$) applies to $E_1$ for all values of $\mu$, to $E_2$ if $\mu\neq\mu^*$, to $E_3$ provided $\mu\neq 1-\mu^*$ and ($m=2$) to $E_4$, $E_5$ for all values of $\mu$. $E_1$, $E_2$, $E_3$ are of positive general type, since none of the linear commensurability relations II of (13) is fulfilled.

THEOREM 11. *In a suitably restricted neighborhood of an $E_i$ $(i=1, 2, 3)$ of negative general type the locus of points which lie on positively [negatively] stable motions having $E_i$ as a unique $\omega$- [$\alpha$-] limiting point is an analytic hypersurface [curve]. In a suitably restricted neighborhood of an $E_i$ $(i=4, 5)$ of negative [positive] general type the locus of points which lie on positively [negatively] stable motions having $E_i$ for a unique $\omega$- [$\alpha$-] limiting point is an analytic two-dimensional manifold.*

The number of classes of stable motions may now be reduced from nine to six.

THEOREM 12. *Excluding the values $\mu^*$, $1-\mu^*$ of $\mu$, if stable motions other than equilibrium motions exist, the $\alpha$-limiting point is $E_4$ or $E_5$ and the $\omega$-limiting point is one of $E_1$, $E_2$, $E_3$.*

To prove the theorem we show that there are no stable motions in the classes (ii), (iii) of Theorem 10.

The motions in the phase space for which the infinitesimal mass remains on the line joining the finite masses lie in the $(x, p)$-plane and are solutions of the differential system

$$\dot{x} = p, \qquad \dot{p} = \Omega_x - p/2^{1/2}. \tag{28}$$

The flow in the $(x, p)$-plane has three equilibrium motions $(a_i, 0)$ $(i=1, 2, 3)$ separated by the points $(-\mu, 0)$, $(1-\mu, 0)$ corresponding to the finite masses. One characteristic constant is positive and the other negative at each equilibrium motion. It follows from Theorem 6 that the locus of points in a sufficiently small neighborhood of $(a_i, 0)$ lying upon motions of (28) having $(a_i, 0)$ as a unique $\alpha$-limiting point is an analytic curve([17]) through $(a_i, 0)$. These equilibrium motions appear in the phase space of (26) as $E_i$ $(i=1, 2, 3)$, and since the motions of (26) having $E_i$ as a unique $\alpha$-limiting point are confined to an analytic curve([17]) through $E_i$, they must lie entirely in the $(x, p)$-plane.

A stable motion of (26) with one $E_1$, $E_2$, $E_3$ as $\alpha$-limiting point and another as $\omega$-limiting point is therefore impossible, for such a stable motion would require the infinitesimal mass to collide with one of the finite masses.

4. **The existence of stable motions.** For Theorem 12 to have content a demonstration of the existence of stable motions other than equilibrium motions is essential. We shall show that two stable motions exist when $\mu=1/2$. Whether other stable motions exist for $\mu=1/2$, or whether any stable motions exist when $\mu\neq 1/2$, are unsolved problems.

Setting $\mu=1/2$, $x=p=0$ in (26) we obtain a differential system of the second order

$$\dot{y} = q, \qquad \dot{q} = \phi(y) - q/2^{1/2}, \qquad \phi(y) = y[1 - ((1/4) + y^2)^{-3/2}],$$

for the motions in the phase space when the infinitesimal mass is restricted to lie on the perpendicular bisector of the line segment connecting the two finite masses.

Since $\phi$ is an odd function of $y$, the motions in the $(y, q)$-plane are paired, the members of a pair being symmetric to each other with respect to the origin. The graph $G$ of $q=2^{1/2}\phi$ is indicated by $ABCOC'B'A'$ in Figure 1. It rises monotonely along $ABC$, $C'B'A'$ and falls monotonely along $COC'$. The flow proceeds to the right [left] in the upper [lower] half-plane, being vertical on the $y$-axis, and is directed downwards [upwards] in the region above [un-

([17]) The curve has no multiple point at the equilibrium motion.

derneath] $G$, being horizontal on $G$, except for the equilibrium motions $B$, $O$, $B'$ corresponding to $E_3$, $E_1$, $E_4$.

The characteristic exponents of $O$ are congugate complex numbers with negative real parts. It follows that, sufficiently near to $O$, the motions have a spiral character([18]) about $O$, that is, as $t \to +\infty$ the point $P$ on the motions tends towards $O$ with the angle between $OP$ and the $y$-axis increasing indefinitely.

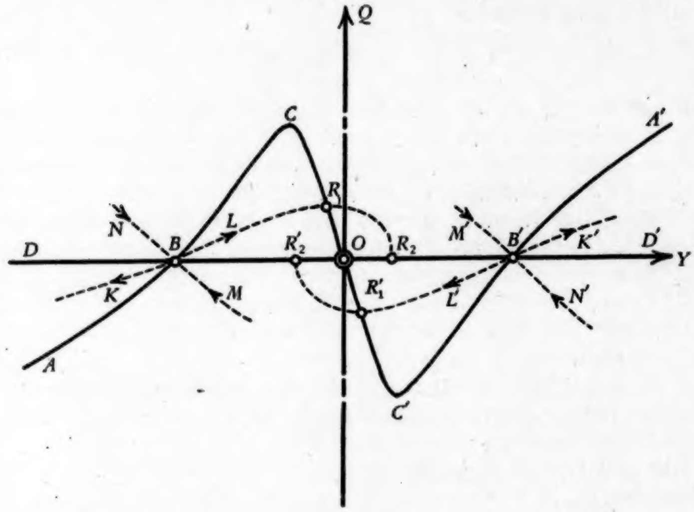


FIG. 1.

At $B'$ one characteristic exponent is positive and the other is negative. Through $B'$ there accordingly pass two analytic arcs $K'BL'$, $M'BN'$, the loci of points lying on motions having $B'$ for unique $\alpha$-, $\omega$-limiting point, respectively. Upon calculation it is found that these arcs are disposed with respect to $G$ as shown in Figure 1.

Consider the arc $B'L'$ which in the immediate neighborhood of $B'$ lies in the open region $B'OC'B'$. In this region the flow is directed to the left and downwards, and since the region contains no equilibrium motion upon which the arc could end, it leaves the region by way of a point $R_1'$ on the arc $OC'$ to enter the open region in the lower half-plane under $G$. Here the flow is upwards and to the left and, since the region contains no equilibrium points, the arc $B'L'R_1'$ must leave it by: (a) the open arc $R_1'O$, (b) $O$, (c) the open segment $OB$ of the $y$-axis, (d) $B$, (e) the open arc $BA$. Clearly (a) is impossible, since on $R_1'O$ the flow is horizontal and directed from right to left. We may

([18]) See, for example, L. Bierberbach, *Differentialgleichungen*, Berlin, 1930, pp. 104–105.

exclude (b) in view of the spiral character of the flow about $O$. The possibility (c) is illustrated in Figure 1 by $B'L'R_1'R_2'$. Paired with such a motion, there is a motion on the arc $BLR_1R_2$ symmetric to it with respect to $O$. The two arcs $BLR_1R_2$, $B'L'R_1'R_2'$ and the segments $BR_2'$, $B'R_2$ of the $y$-axis enclose a region containing $O$ into which the motion on $B'L'R_1'R_2'$ enters and never leaves. It cannot leave by way of the open segments $BR_2'$, $B'R_2$, for on the former the flow is vertically upwards and on the latter vertically downwards. Departure by way of $B$ or $B'$ is ruled out by Theorem 10 and it cannot meet either $BLR_1R_2$ or $B'L'R_1'R_2'$ for such a point of meeting can occur only at an equilibrium motion. The motion accordingly remains in this region and is therefore stable. Paired with it, there exists a second stable motion. (d) is impossible by Theorem 10.

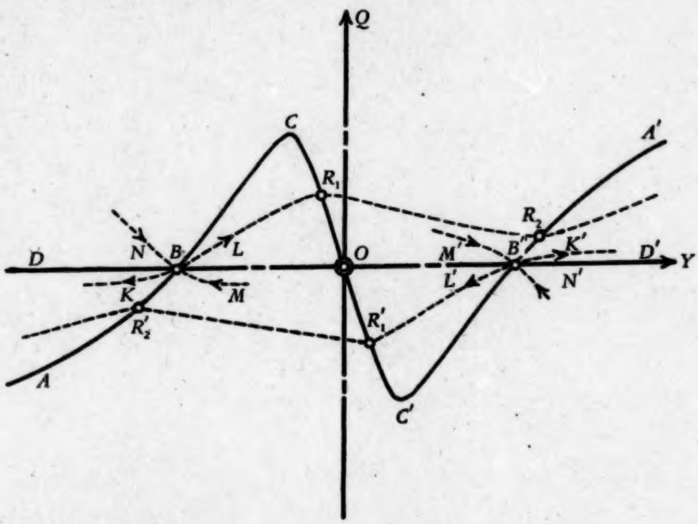


FIG. 2.

The theorem is accordingly proved, provided we can rule out the possibility (e) pictured in Figure 2 by $B'L'R_1'R_2'$. Paired with a motion on this arc, there is a motion symmetric to it with respect to $O$ on the arc $BLR_1R_2$. These two arcs taken with $B'R_2$, $BR_2'$ bound a region containing $O$ and the motion $M'B'$ in its entirety. The motion $M'B'$ would possess $O$ or $B$ as $\alpha$-limiting point and $B'$ as $\omega$-limiting point, thus contradicting Theorem 10. Hence (e) is impossible.

## VI. THE RESTRICTED PROBLEMS OF ELLIPTIC AND HYPERBOLIC TYPE

1. **The differential equations of motion.** If the mass $\mu$ moves in an elliptic

[hyperbolic] orbit about the mass $1-\mu$, it is known[19] that

$$\rho = a(1 - e\cos\phi), \qquad a^{3/2}(\phi - e\sin\phi) = \kappa\tau,$$
$$[\rho = a(e\cosh\psi - 1), \qquad a^{3/2}(e\sinh\psi - \psi) = \kappa\tau].$$

The independent variable $\bar{t}$ in IV turns out to be proportional to $\phi$ $[\psi]$ and we find

$$\rho = a(1 - e\cos\bar{t}/a^{1/2}), \qquad [\rho = a(e\cosh(\bar{t}/a^{1/2}) - 1)],$$

so that

$$l = \log a(1 - e\cos\bar{t}/a^{1/2}), \qquad [l = \log a(e\cosh(\bar{t}/a^{1/2}) - 1)],$$
$$k = \frac{e\sin\bar{t}/a^{1/2}}{a^{1/2}(1 - e\cos\bar{t}/a^{1/2})}, \qquad \left[k = \frac{e\sinh\bar{t}/a^{1/2}}{a^{1/2}(e\cosh(\bar{t}/a^{1/2}) - 1)}\right],$$

are to be taken in the canonical equations (25) with $H$ defined as in (18) and the dot denoting differentiation with respect to $\bar{t}$.

For rectilinear problems the independent variable $t$ in IV is employed to yield

$$\begin{aligned} &\rho = 2a\,\mathrm{sech}^2\, t/2^{1/2}, && [\rho = 2a\,\mathrm{csch}^2\, t/2^{1/2}], \\ &k = -(1/2^{1/2})\tanh t/2^{1/2}, && [k = -(1/2^{1/2})\coth t/2^{1/2}]. \end{aligned} \tag{29}$$

The differential equations (1) take the simple form

$$\ddot{x} + k\dot{x} = \Omega_x, \qquad \ddot{y} + k\dot{y} = \Omega_y.$$

It is readily verified that the differential equations for the two types interchange when $t$ is replaced by $t+(\pi/2^{1/2})i$. One obtains the canonical form (25) for these equations by taking the above values of $k$ and placing $b=0$ in the definition of $H$ in (16).

2. **Limiting motions.** In the general problem of hyperbolic type $e>1$ and the $\omega$-limiting motions are the motions

$$\begin{aligned} &x = x_0 + a^{1/2}p_0(1 - e^{-\bar{t}/a^{1/2}}), && y = y_0 + q_0 a^{1/2}(1 - e^{-\bar{t}/a^{1/2}}), \\ &p = p_0 e^{-\bar{t}/a^{1/2}}, && q = q_0 e^{-\bar{t}/a^{1/2}}. \end{aligned} \tag{30}$$

The $\omega$-limiting motions of a positively stable motion are equilibrium motions in (30). However since these equilibrium motions do not form a discrete set, it cannot be assumed that the motion tends uniformly towards a definite equilibrium motion, for it is conceivable that the motion tends toward a set of equilibrium motions.

For rectilinear problems of elliptic or hyperbolic type, it follows from (29) that the $\alpha$- and $\omega$-limiting motions are, respectively, motions of the dissipative and acquisitive rectilinear problems of parabolic type. A positively [negatively] stable motion approaches one of the equilibrium motions $E_i$ of the rectilinear problem of parabolic type uniformly as $t\to+\infty$ $[-\infty]$. It will be observed that as $t\to\pm\infty$ the finite masses tend towards collision.

(19) See F. R. Moulton, op. cit., pp. 158–159 and pp. 177–178.

University of Maryland,
College Park, Md.